

Predicting Employer Health-Insurance Renewals

A synthesis of the standard trend indices, the five-year economic outlook, the ACA rate-filing signal, a quantitative out-of-sample backtest of forecasting rules, and a specialty-drug leading-indicator model — assembled into a single recommended renewal-prediction architecture for US commercial group medical.

Scope

US commercial group medical — small group (1–50/100) & large group (50/100+)

Basis

Four sourced research briefs, one 26-year backtest, one specialty-drug model

Status

Internal — pre-client review. Vendor figures require re-pull & license.

Prepared for

Sam Newland, Benefitra

Prepared by

Benefitra Actuarial Intelligence

Date

September 2026

A note on figures. Every quantitative claim in this paper carries a provenance label: **SOURCED** (verified against a cited URL during research), **ESTIMATE** (an analyst blend or modeled figure, not a single published number), and **RE-PULL** (a proprietary vendor figure referenced from research but requiring a licensed re-pull before any client-facing use). See Section 10 for the full key.

ABSTRACT

There is no single “renewal index,” and the search for one is the wrong project. The right project is a small, disciplined system that reads what is knowable and forecasts what is not.

Employer health-insurance renewals resist a one-number forecast because two distinctions govern the entire problem. First, *forward trend* (the benefits-held-constant number carriers price on) is not the same as *realized premium growth* (what employers pay after they buy coverage down); the first predicts, the second calibrates. Second, small-group business is adjusted-community-rated and *rate-filed with each state months ahead* — its next-year price can be read rather than forecast — whereas large-group business is experience-rated with no public filing and must be projected from trend surveys and the group’s own claims.

This paper synthesizes five workstreams: (i) a ranked landscape of national and per-state trend indices; (ii) the “GDP + excess” architecture that health economists use for the five-year outlook, yielding a national glide from roughly 7.5% near-term toward 5–5.5% long-run; (iii) the ACA requested-versus-approved rate-filing signal and its regime-gated limits; (iv) a 26-year out-of-sample backtest establishing that a light exponential smoother (EWMA, $\alpha \approx 0.5$) beats both a naive last-year carry and a pure long-run mean; and (v) a specialty-drug leading-indicator model that separates one-time *cure pulses* (Hepatitis C) from persistent *chronic waves* (GLP-1). We conclude with a concrete two-engine build specification — a small-group filings reader and a large-group trend-blend engine — over a Getzen five-year glide, a smoothed-recent base rule, a named-basis-point specialty-drug overlay, and Benefitra’s own placed-case renewal book as the proprietary calibration index that no competitor can replicate.

Key findings

1. **No single index exists, and the best predictor differs by block.** Small-group next-year prices are *filed* and should be read from ACA rate filings; large-group must be forecast from a survey blend and the group’s own claims. Treat them as two engines.
2. **Next year is running hot.** The live-sourced anchor for group commercial trend is **~9%** (PwC 2027 Group medical, “highest in 17 years”). Realized 2025 employer premiums rose +5%/+6% single/family after buy-downs. Price is tame; utilization and specialty pharmacy are the drivers.
3. **The five-year path bends down, not up.** The “GDP + excess” architecture gives a glide from ~7.5% near-term to ~5–5.5% long-run — a national five-year average of **~6.5–7.0%** (estimate), ± 1 pt/yr, with ± 1 –1.5 pt of state spread.
4. **A smoother beats the obvious alternatives.** Across 18 rolling out-of-sample test years, an EWMA with $\alpha \approx 0.5$ (roughly $0.6 \times \text{last} + 0.4 \times \text{prior}$) had the lowest error — better than naive last-year (which over-reacts to spikes) and far better than a pure long-run mean (worst).
5. **Specialty drugs are a renewal shock you can see coming.** The regulatory and pipeline calendar makes launches visible 1–3 years ahead; the drug’s biology tells you the shape. A *cure* is a pulse that decays (Hep C fell ~80% from its 2015 peak by 2018); a *chronic* therapy is a wave that plateaus (GLP-1). Model each as a named basis-point line item, kept separate so a pulse can be removed as it decays.
6. **The differentiating edge is a pharma-pipeline overlay plus your own book.** The specialty/GLP-1 pipeline is the longest-lead public signal (12–36 months). Layered on the standard blend and calibrated against Benefitra’s placed-case renewals, PBM, and stop-loss files, it is the “standard + X” composite worth building.
7. **The requested-versus-approved gap is real but narrow.** The *requested* trend leads; the *gap* only carries information in prior-approval states, and large group has no filed data at all. Use the approved single-risk-pool change as a Bayesian anchor for small group.

CONTENTS

1. **Introduction — why renewal prediction is hard.** The two governing distinctions and the rating-line nuance.

2. **The landscape of renewal indices.** National and per-state, small versus large group, ranked.
3. **What economists use & the five-year outlook.** CMS/CBO/Getzen "GDP + excess"; the glide path.
4. **The requested-versus-approved rate-filing signal.** Machinery, regime-gating, and verdict.
5. **A quantitative backtest of forecasting rules.** Which historical lookback best predicts next year.
6. **Specialty drugs as leading indicators.** The pulse-versus-wave framework and model.
7. **Toward a better composite.** "Standard trend + a differentiating factor."
8. **A recommended renewal-prediction architecture.** The two-engine build specification.
9. **Does an indexed fully-insured renewal change the recommendation?** A 35-segment impact test of the FI-renewal assumption.
10. **Limitations & data provenance.** The honesty notes, consolidated.
11. **References.** Grouped source list.

1 Introduction — why renewal prediction is hard

Two distinctions decide which data source can even answer the question. Get them right before choosing an index.

An employer's renewal is, on its face, a single number: the percentage by which next year's premium (or, for a self-funded plan, next year's expected claims plus fixed costs) exceeds this year's. The temptation is to look for the one index that predicts it. That temptation should be resisted. The renewal number is the output of a chain — unit price, utilization, service intensity and mix, plan-design leverage, demographic drift, area cost, carrier margin, and regulatory action — and different links in that chain are observable through entirely different instruments depending on how the group is rated. Before any index is chosen, two distinctions must be settled.

Distinction 1 — Forward "trend" (to predict) versus realized premium growth (to calibrate)

Trend is the forward-looking, benefits-held-constant number carriers price on: price × utilization × mix, before any plan change. This is what PwC, Segal, Aon, and WTW forecast.^{1, 14, 15} **Realized premium growth** is what KFF and Mercer measure *after* employers change plans and buy coverage down.^{2, 15} The two are not interchangeable: realized growth typically runs 1–2 points below trend precisely because employers purchase the difference through higher deductibles, narrower networks, or richer cost-sharing. **Use trend to predict the renewal; use realized to calibrate the model against what employers actually paid.** Confusing the two is the classic error.

Distinction 2 — Small group is rate-filed; large group is experience-rated

Small group is adjusted-community-rated and **rate-filed with the state**: the next-year price is literally submitted months ahead through SERFF and the federal rate-review system.^{9, 10} You *read* it; you do not forecast it. **Large group** is experience-rated with no public filing, so forward trend surveys plus the group's own claims are the only signals available.^{1, 14, 15} This single split determines the entire modeling architecture in Section 8: two engines, not one.

The rating-line nuance: ALE, small group, and where the boundary actually sits

The small-versus-large boundary is not uniform. Under the ACA, the *small-group market* is defined federally as employers with 1–50 full-time-equivalent employees, though states may expand it to 1–100 (California, Colorado, New York, and Vermont historically did). Adjusted community rating — where premiums may vary only by age (3:1 band), geography, tobacco use, and family size, but not by health status — applies within that market, and rates are filed. Separately, the *Applicable Large Employer* (ALE) threshold of 50 full-time-plus-FTE employees governs the employer mandate and reporting, not the rating method. An employer can therefore be an ALE for mandate purposes while still being rated in the small-group (community-rated, filed) market in a 1–100 state, or can sit just above the small-group line and be experience-rated. For renewal prediction the operative question is always the *rating method*, not the mandate status: is this group's price community-rated and filed, or is it experience-rated and negotiated? The answer selects the engine.

The right question is never "what is the renewal index?" It is "for this group, is the next-year price filed and readable, or unfiled and to be forecast?"

With those distinctions in place, the remainder of this paper proceeds from the observable to the modeled: the standard indices (Section 2), the long-run economic architecture (Section 3), the filing signal (Section 4), the empirical question of which lookback predicts best (Section 5), the specialty-drug overlay (Section 6), the composite design (Section 7), and the build specification (Section 8).

2 The landscape of renewal indices

National and per-state, small versus large group. What actually predicts the renewal 6–12 months out.

A dozen organizations publish something they call a medical trend, cost index, or premium survey. They are not equally useful for prediction, and several that are widely cited (CPI-Medical, the Milliman Medical Index level) are among the least useful. This section ranks them by their power to predict a *next-year* renewal, holding the two distinctions of Section 1 in mind: for small group the filed number dominates; for large group a survey blend plus own claims dominates.

The live-sourced anchors

9.0%

PwC 2027 **Group** medical trend — "highest in 17 years." Individual 8.5%; 2026 restated up to 9.0%/8.5%.

SOURCED · PwC Behind the Numbers 2027¹

+5% / +6%

Realized 2025 employer premiums, single / family (single \$9,325; family \$26,993). Small-firm family \$26,054 vs large \$27,280.

SOURCED · KFF EHBS 2025²

+5.2%

CPI hospital services, Jul 2025 → Jul 2026. Physician only +2.4%; medical commodities ~2.7% — price is not the driver.

SOURCED · BLS CPI³

The gap between the ~9% forward trend¹ and the +5–6% realized premiums² is Distinction 1 in a single frame: employers absorbed roughly three points through plan changes. The gap between the ~9% trend and the +2.4–5.2% CPI medical components³ is the utilization-and-intensity story: pure provider price is running near general inflation, so the excess is coming from how much care is used and how intensive it is — principally specialty pharmacy, behavioral health (PwC cites +62% utilization since 2018), and new provider revenue tools.¹

Small group — ranked

#	Index / source	Why it ranks here	Label
1	State ACA small-group rate filings (SERFF / CMS URRT projected trend)	The actual filed price change, per state, forward. Read it, do not forecast it. The crown jewel.	SOURCED
2	Segal Health Plan Cost Trend Survey	Independent forward trend with strong component detail (medical ~8%+, Rx ~11–12%, specialty Rx ~15%+).	RE-PULL
3	KFF EHBS small-firm (10–199)	Realized, not forward — use for calibration, not prediction.	SOURCED
4	PwC Group (national overlay)	Directional national anchor to blend over the filed base.	RE-PULL
–	<i>Weak</i> : Milliman Medical Index level, CPI-Medical	Lagging or definitionally noisy for a next-year renewal.	

Table 1. Small-group next-year predictors, ranked by forward predictive power. The crown jewel — ACA rate filings — is the one source nobody markets as an "index."

Large group — ranked

#	Index / source	Why it ranks here	Label
1	Segal + PwC (Group) + Aon blend	The trend baseline carriers apply; methodologically independent → blend them. Currently ~9%.	RE-PULL
2	WTW + Mercer projection	Second-opinion forward trend; Mercer also gives realized (post-plan-change, ~5.5–6%).	RE-PULL
3	Group's OWN claims experience	Dominates every other signal for any specific account — the single most important input at the case level.	OWNED
4	KFF / Mercer realized	Validation layer against actual paid premiums.	SOURCED
–	Weak: CPI / PCE, Milliman Medical Index level	Macro price series lag commercial trend badly.	

Table 2. Large-group next-year predictors. No filings exist, so a survey blend plus the account's own claims carries the forecast.

The per-state signal

Small group / individual. CMS Rate Review Public Use Files and RateReview.HealthCare.gov give per-insurer proposed and approved rate changes plus the URRT projected trend (unit cost × utilization) for every state.^{9,10} Weighting each carrier's filed change by covered lives or market share produces a per-state small-group forecast. It is machine-ingestible with effort and refreshes every summer as filings publish.

Large group. No filings exist, so the national trend blend is flexed by geographic **area factors** — Milliman Health Cost Guidelines if licensed,¹⁶ otherwise BLS metro CPI/PPI plus CMS Geographic Practice Cost Indices and the hospital wage index as a proxy. Benefitra's own placed-case renewal book is the highest-value *owned* per-state large-group signal, and no competitor can replicate it.

Best single national point estimate for next year: blend PwC Group, Segal medical, and Aon — three methodologically independent surveys — to arrive at **~9% group commercial** ESTIMATE FROM RE-PULL INPUTS. The single most important, most decision-useful, per-state, machine-ingestible source is the one nobody markets as an index: the ACA small-group rate filing. It should be the foundation of the small-group book.

3 What economists use & the five-year outlook

Multi-year forecasting is not an index — it is a decomposition. All three official models share one architecture.

Where the one-year question is answered by trend surveys and filings, the five-year question is answered by a different discipline entirely. There is no five-year renewal index, and no one credible publishes one. Instead, health economists decompose nominal medical trend into three additive components and force the long run to converge toward the growth of the economy:

```
nominal medical trend = general inflation + real per-capita income growth + "excess" growth
where excess = the growth in health care's share of GDP (technology & intensity)
```

This "GDP + excess" architecture is the shared backbone of every serious multi-year model. The reason it matters commercially is that it disciplines the long-run tail: health spending cannot outgrow GDP forever, so any credible five-year forecast must bend the near-term rate down toward the economy's growth rate rather than extrapolating a hot recent year straight out.

The four instruments

Model	Horizon	Core primitive	Current read
CMS NHE projections ⁴	10-yr, annual (2025–34)	Total NHE +5.4%/yr vs GDP +4.1% → excess ~1.3pp; health share 18.0% → 20.6%	Private health insurance +5.0%/yr <i>aggregate (a floor)</i>
CBO (Budget Outlook + LTBO) ⁵	10–30 yr	“Excess cost growth” wedge (per-capita health – per-capita GDP), declines toward ~1%	CPI +2.2%, GDP deflator +2.0%, real per-capita income +1.3–1.4%
Getzen / SOA model ⁶	Long-run (10+ yr)	%trend = %income + %inflation + %excess; steady state = GDP+0%	Long-run ~5.0%/yr (2025 update)
Altarum HSEI ⁷	Monthly	Price-versus-utilization split (nearest thing to a leading index)	Growth is utilization-led, so trend is sticky even as CPI cools

Table 3. The economists’ toolkit for multi-year medical cost forecasting. All share the “GDP + excess” decomposition; they differ in horizon and in what they hold as the core primitive.

The Getzen model, explained

The Getzen / SOA Long-Run Medical Cost Trend Model — developed by Prof. Thomas Getzen of Temple and maintained by the Society of Actuaries as a free, annually-updated Excel tool — is the standard actuarial instrument for the long horizon.⁶ It is explicitly *not* a short-term year-by-year forecaster. Its structure is a glide in three zones:

- **Years 1–4** are user-modifiable short-term inputs — you supply the near-term trend from surveys and your own book.
- **Years 5–9** are a linear transition.
- **Years 10+** resolve to per-capita income growth + inflation + an excess term (GDP + X), with the steady state at GDP + 0% — the point at which health’s share of GDP stops rising.

Two stabilizers keep the tail honest: a *share-resistance point* (a 2025 baseline of 0.180, i.e. an 18.0% health share of GDP, past which excess growth is damped) and a *year-limit* (around 2075, by which excess is forced to zero). The 2025 baseline parameters — inflation +2.6%, real per-capita income +1.4%, long-run excess +0.9% — sum to a **long-run medical trend of ~5.0%/yr**, declining thereafter toward income growth. Getzen is candid that there is *no usable “medical inflation” index*: CPI-Medical and PPI are “almost useless” because rebates, subsidies, and confidential contracting obscure true price. Expenditure — spend as a share of GDP — is the only well-defined quantity, so the model anchors on **shares, not price indexes**.⁶

The recommended five-year blend

Combining the four instruments through the Getzen glide architecture yields a four-step recipe:

1. **Anchor the long run** at Getzen ~5.0% (years 5+).⁶
2. **Set the near term** (years 1–2) at ~7–8% gross per-member (Mercer/Aon/WTW/PwC), which is higher than the CMS aggregate because that aggregate is diluted by falling private-insurance enrollment — see the caveat below.^{1, 4, 15}
3. **Glide between them** along the Getzen transition.
4. **Flex by state** ±1–1.5 pt via each state’s historical CMS State Health Expenditure spend-per-capita growth relative to the national rate.

The result is a **national five-year average renewal of ~6.5–7.0%/yr** ESTIMATE for a stable, un-bought-down plan (net of plan changes running ~1–2 pts lower), presented with a ±1 pt/yr uncertainty band. The central line in Figure 1 is an estimate built on sourced CMS/Getzen/CBO parameters; only the exact yearly points are modeled, not published.

Recommended 5-Year National Renewal Glide Path

Getzen "GDP + excess" architecture: near-term ~7.5% mean-reverting toward ~5-5.5% long-run. ESTIMATE, un-bought-down plan.

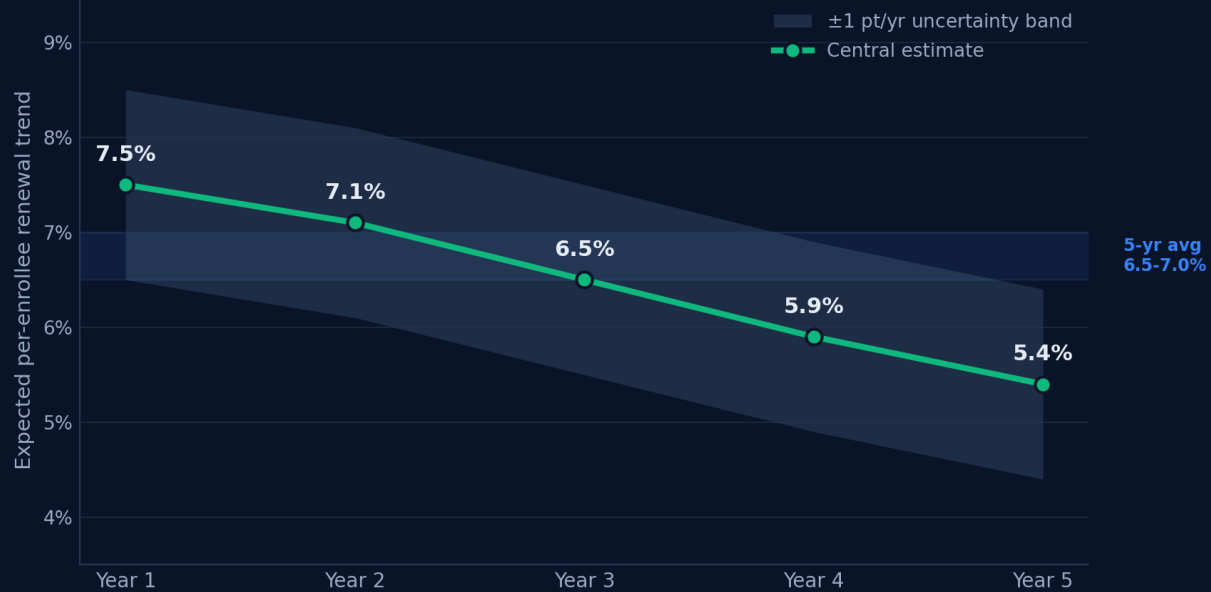


Figure 1. Recommended five-year national renewal glide path. Getzen "GDP + excess" architecture: near-term ~7.5% mean-reverting toward ~5-5.5% long-run, for a five-year average of ~6.5-7.0%. The central line is an **ESTIMATE** built on sourced CMS/Getzen/CBO parameters; the band is ± 1 pt/yr forecast uncertainty (Getzen accuracy $\sim \pm 1\%$ /yr, compounding). Per-enrollee, un-bought-down plan.

The per-enrollee-versus-aggregate caveat — do not misread the CMS 5.0%

The CMS private-health-insurance projection of +5.0%/yr is an *aggregate* figure, depressed by *falling* private-insurance enrollment over the projection window. The *per-enrollee* number — which is the renewal concept — runs meaningfully higher, and CMS says so explicitly.⁴ Treat aggregate PHI growth as a floor, never as the renewal trend. This is why the near-term anchor in step 2 sits at ~7-8% (per-member surveys) rather than at the ~5% aggregate: they measure different things.

One further empirical note from Altarum's monthly data reinforces the shape: recent growth has been *utilization-led*, not price-led (the Health Care Price Index has run near +3%, while utilization has run near +4% and Rx use notably higher).⁷ Utilization-driven trend is stickier than price-driven trend, because it does not fall when general inflation cools. That is the mechanical reason the near-term stays hot even as CPI-Medical looks tame — and the reason the specialty-drug overlay of Section 6 matters so much.

4 The requested-versus-approved rate-filing signal

A genuine leading signal — but a narrow one, and it does not exist at all for large group.

Because small-group prices are filed (Distinction 2), a natural question is whether the *gap* between what carriers *request* and what regulators *approve* carries predictive information. It does, but less than one might hope, and its value is concentrated in a handful of states.

The machinery

Under ACA §2794, the annual rate-review process requires individual and small-group carriers to file the Unified Rate Review Template (URRT): Part I data, Part II written justification, and Part III actuarial memorandum.^{9,10} The federal review threshold is 15% (it was 10% from 2011-2013); filings at or above it draw heightened scrutiny. Critically, **large group is generally not federally rate-reviewed**: filing requirements vary by state law and there is no national public dataset. That is a hard limit on large-group modeling, not a gap to be closed.

Data sources (individual + small group, per state × carrier × year)

- **CMS Single-Risk-Pool "Rate Review" PUF** (2014–2026, bulk CSV) — requested changes (URRT Part I) plus a review-status field.⁹
- **RateReview.HealthCare.gov** — requested plus *final approved* justifications, per state and issuer (a web application, not clean bulk data).¹⁰
- **Peterson-KFF Health System Tracker** annual premium brief — individual/marketplace, national and per-state issuer-level, requested and approved, analysis-ready. The 2026 brief (updated 15 Jan 2026) reported an approved median of 18% (range −10% to +59%, IQR 12–27%); the 2025 brief reported a 7% requested median.¹¹
- **Georgetown CHIR** — deep actuarial-memorandum reads (2026: 178 filings across 28 states plus DC).¹²
- **State DOI dashboards / SERFF** — some show requested versus approved (CA, NY, OR, CO, RI, WA); large group only where state law requires it.

The regime distinction that governs whether the gap is real

Whether the requested-versus-approved gap carries any information depends entirely on the state's rate-review regime. In **prior-approval** states a regulator can cut or reject a filing, so the gap encodes a real regulatory haircut. In **file-and-use** or **use-and-file** states the requested rate is effectively the rate, so the gap is approximately zero by construction. Using the gap therefore requires a per-state prior-approval-versus-file-and-use classification table (buildable from the KFF "effective rate review" indicator plus DOI statutes) — a gap to pull, not something retrieved live in this research.

Verdict

- **Requested trend is the stronger leading signal.** Published each spring/summer for the following January, it encodes the carrier's forward view of trend, drug and GLP-1 loadings, utilization, and policy shocks — for 2026, filings disclosed effects such as the expiration of enhanced premium tax credits (~+4 pts) and tariffs (~+3 pts).^{11,12}
- **The gap itself is weaker and situational.** Approved rates historically track requested rates closely — regulators trim at the margin, and large cuts occur only in prior-approval states — so the gap is usually small and adds limited standalone lift. Its value concentrates in CA, NY, OR, CO, and RI.
- **Filed trend predicts directionally, not deterministically.** It sets the base or anchor for the small-group block; an individual group deviates via its plan and metal mix, area factor, demographics, and size.

How to use it in a model

Use the **approved single-risk-pool change as a Bayesian prior/anchor** for the small-group block (the requested figure overstates in prior-approval states); then let a specific group's features shift the posterior off that anchor. Treat **requested trend** as the earliest read and the **requested-versus-approved gap** as a second-order feature, interacted with a prior-approval flag. The feature set, per state × carrier × year, is: filed trend (earliest, leading), approved trend (corrected anchor), the gap (regime-gated), a regime categorical, a three-year trailing filed trend and its volatility, and disclosed adjusters from URRT Part II (the ~8% medical-trend assumption for 2026, specialty/GLP-1 loadings, policy shocks).

Hard limits (model card):

- Large group has **no feature here** — it is not filed.
- An approved *block average* is not what any one group sees.
- Value is concentrated in a handful of prior-approval states.
- The model must be **as-of-date aware** — the requested → approved swap must not leak look-ahead information into training.
- The ACA small-group block is shrinking → thin cells in some states.

5 A quantitative backtest — which lookback best predicts next year

If you must forecast from history alone, which rule wins out of sample? We tested fourteen.

Sections 2–4 concern *external* signals. But every model needs a base rule: given only the recent history of a group's (or a block's) trend, what is the best pure-time-series forecast of next year? This is not a rhetorical question — it is testable, and the answer is not the obvious one. We ran a rolling out-of-sample backtest to settle it.

Methodology

The series used is the KFF Employer Health Benefits Survey all-firms family-premium annual increase, 2000–2025 (26 observations).^{2, 24} Recent values were confirmed live (2024 = +7%, 2025 = +6%) and cross-checked against confirmed cumulative figures (2019–24 = +24%, 2020–25 = +26%); interior years are estimates from training knowledge, reconciled against those cumulatives to within rounding **EST (INTERIOR)**. The evaluation is a genuine rolling, out-of-sample exercise: for each test year 2008–2025 (18 years), each candidate rule is fit only on data available up to that point and asked to predict the next year; errors are then aggregated as root-mean-square error (RMSE) and mean absolute error (MAE). Fourteen rules were compared, spanning naive persistence, moving averages, exponential smoothers, fitted autoregressions, and the expanding long-run mean.

The underlying structure

Before ranking rules it helps to understand the series' character. The autocorrelation function decays (0.70, 0.58, 0.37 at lags 1–3) and the partial autocorrelation cuts off after lag 1 — the signature of an **AR(1)** process. Fitting it gives:

$$y_t = 1.80 + 0.703 \cdot y_{t-1} \quad (p < 0.001, \quad R^2 = 0.50)$$

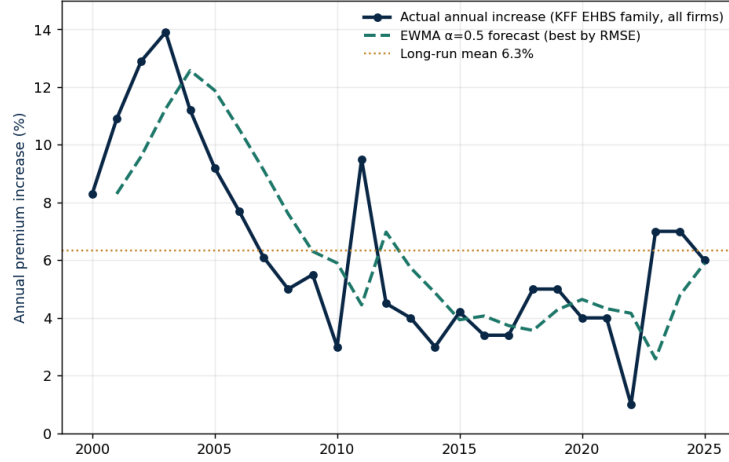
The coefficient of 0.70 — positive but well below 1 — is the whole story in one number: trend has **momentum** (last year matters a lot) but is **mean-reverting** (it pulls back toward a long-run center of ~6.0%). Adding a second lag (AR(2)) improves fit trivially (R^2 0.50 → 0.52). The full-sample optimal weights on the last three years, [0.62, 0.37, 0.01], are effectively an exponential smoother with $\alpha \approx 0.5$ –0.6 — which foreshadows the winner.

Results

#	Predictor	RMSE	MAE	Verdict
1	EWMA, $\alpha = 0.5$	2.263	1.762	Winner — best RMSE
2	Mean of last 3	2.286	1.824	Essentially tied
3	Mean of last 2	2.288	1.786	Essentially tied
4	Mean of last 4	2.293	1.888	Essentially tied
5	Blend 0.3·last + 0.7·mean3	2.297	1.755	Tied
6	EWMA, $\alpha = 0.7$	2.354	1.730	Good
7	Blend 0.5·last + 0.5·mean3	2.357	1.709	Best MAE
8	EWMA, $\alpha = 0.3$	2.361	2.005	Good
9	Mean of last 5	2.399	2.007	Good
10	Optimal weights, last 3 (expanding fit)	2.610	1.976	Overfit
11	Last year (naive / random walk)	2.665	1.761	Poor — over-reacts to spikes
12	AR(1) regression (mean-reverting)	2.738	1.971	Overfit on 20 pts
13	AR(2) regression	3.178	2.233	Overfit
14	Expanding long-run mean	3.615	3.180	Worst — ignores momentum

Table 4. Rolling out-of-sample backtest, 18 test years (2008–2025), ranked by RMSE (lower is better). A light smoother wins; naive last-year is poor; the pure long-run mean is worst. The top nine are statistically bunched.

U.S. small-group / small-firm fully-insured renewal trend, 2000-2025



Spine: KFF Employer Health Benefits Survey family-coverage increase. 2024/25 confirmed live; 2000-2023 [EST] from KFF training-knowledge (live scrape blocked). All-firms proxy — see small-group caveat.

Backtest error ranking: smoothing beats naive-last & pure-mean

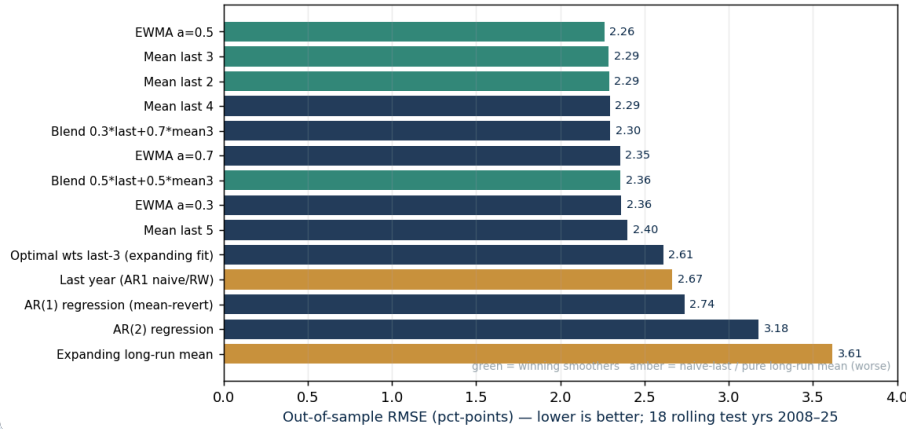


Figure 2. Backtest of forecasting rules. Out-of-sample prediction error across 18 rolling test years. A light exponential smoother (EWMA $\alpha \approx 0.5$) tracks actuals best; naive last-year over-shoots on spike years (note 2011, 2021-22); the expanding long-run mean is flat and misses every turn. Series: KFF all-firms family-premium increase, 2000-2025; interior years **EST**.

Verdict and the practical rule

A smoother wins. The best rule is an EWMA with $\alpha \approx 0.5$ — roughly **next-year $\approx 0.6 \times \text{last} + 0.4 \times \text{prior}$** — or, equivalently, a trailing two-to-three-year average. Naive last-year is poor because it over-reacts to spike years; the pure long-run mean is worst because it ignores momentum entirely; and the fitted AR and “optimal-weight” rules overfit the roughly 20-point training sample. The series is momentum-driven but mean-reverting, so the right posture is to **lean on the recent reading yet pull it toward the ~6% center**.

BASE RULE The recommended time-series base

Forecast next year as an **EWMA with $\alpha = 0.5$** on the recent trend ($\approx 0.6 \times \text{last} + 0.4 \times \text{prior}$), then nudge ~10-20% toward the ~6% long-run mean when the recent reading is far from center. Hand version: blend last year with the trailing-three-year average at roughly 40/60. **Not** naive last-year; **not** the pure long-run mean.

Honest caveats

There are only 18 out-of-sample test points, and the top nine rules are bunched within an RMSE band of 2.26–2.40 — they are *not* statistically separable within the smoother pack. The robust, defensible claim is therefore the weaker one: **any reasonable smoothing beats both naive last-year and the pure long-run mean**. The exact winner (EWMA 0.5 versus mean-of-3) is a coin-flip on this sample. The series is also a proxy — all-firms family premium, which is large-/self-funded weighted — not a literal small-group-rated book; the persistence *structure* transfers, but the *level* and volatility of a real small-group book will be higher, which if anything *strengthens* the case for smoothing. Interior years are estimates **EST** pending a citation-grade re-pull of the full KFF series.

6 Specialty drugs as leading indicators

A new specialty drug is a renewal shock you can see coming — and its biology tells you the shape.

Of all the added factors beyond standard trend, the specialty-drug pipeline offers the longest genuine lead and the cleanest public data. The reason is structural: the FDA and clinical-trial calendar make a launch visible months to years ahead, and the drug's *mechanism* tells you whether the cost hit will be a one-time **pulse** that decays or a persistent **wave** that plateaus. Hepatitis C is the pulse archetype; GLP-1 obesity therapy is the wave archetype. The two shapes demand opposite treatment in a renewal model, and confusing them is expensive in both directions.

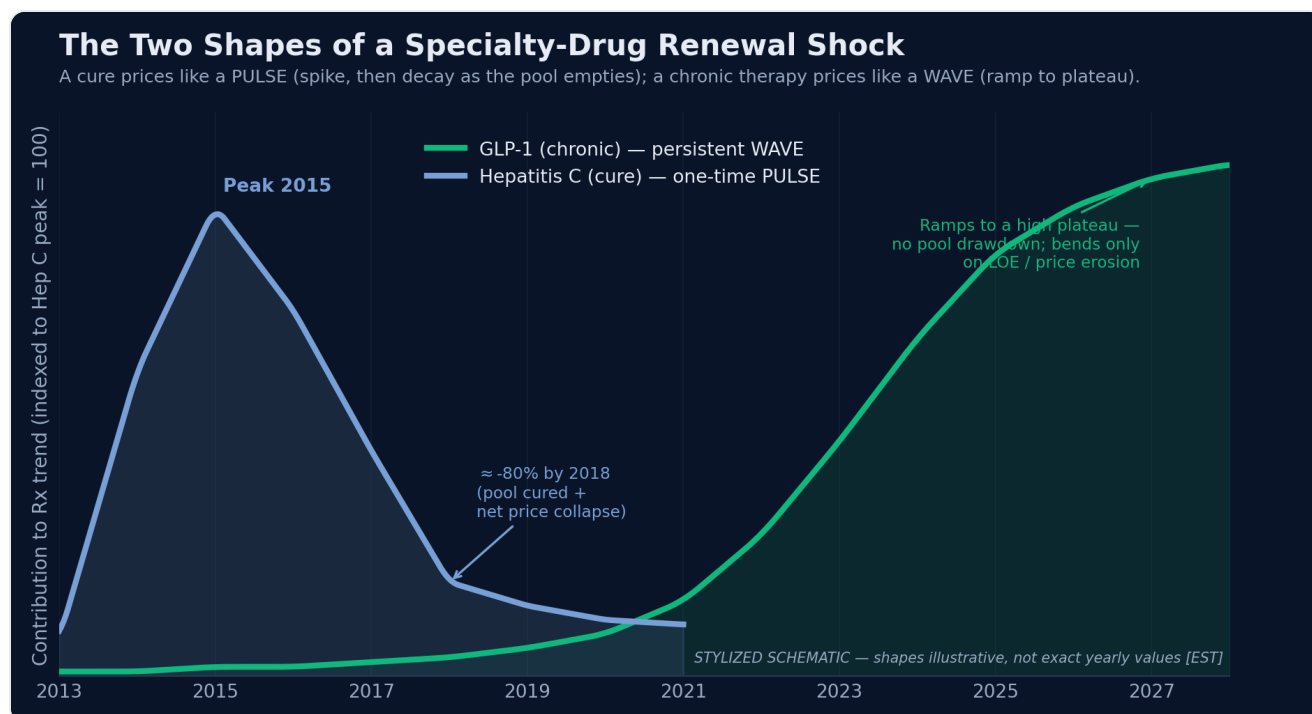


Figure 3. The two shapes of a specialty-drug renewal shock. A cure (Hep C) prices like a pulse: a ~2-year ramp, a 2015 peak, then a ~80% decay by 2018 as the prevalent pool is cured and competition collapses net price. A *chronic* therapy (GLP-1) prices like a wave: it ramps toward a high plateau and bends only on loss of exclusivity or price erosion, never on pool exhaustion. **Stylized schematic** — shapes are illustrative and calibrated to the documented arcs; exact yearly values are **EST**.

Case 1 — Hepatitis C: the pulse-then-amortize archetype

The spike. Gilead's Sovaldi (sofosbuvir) was FDA-approved in December 2013 at a list price of \$84,000 per 12-week course (~\$1,000/pill); Harvoni followed in October 2014 at ~\$94,500.¹⁸ Peterson-KFF explicitly flags these two drugs as *the* exception to otherwise-modest 2010s drug-spend growth.¹¹ Retail Rx spending grew ~12.2% in 2014 — the fastest since 2002, driven largely by Hep C **EST**;⁴ Express Scripts reported Hep C spend up 742% year-over-year as the single largest driver of total drug spend **EST**;²⁰ and Gilead's total revenue jumped from \$24.9B (2014) to a \$32.6B peak (2015) almost entirely on the franchise.¹⁹

The decay — why the pulse self-limits. Three mechanisms drive the ~80% fall from peak:

1. **It is a cure, not a chronic therapy.** One ~12-week course clears the virus, so each treated patient *leaves* the eligible pool. Gilead’s own disclosures attribute post-2017 revenue drops not only to pricing pressure but to “the number of suitable patients decreased.”¹⁹ That is the prevalent-pool drawdown in one sentence.
2. **Net-price collapse from competition.** AbbVie’s Viekira Pak (Dec 2014) and then Mavyret (Aug 2017, ~\$26,400 list) triggered formulary wars; Gilead announced a 46% average discount off list in 2015, and the net course cost fell to ~\$64,700 by 2020.¹⁸
3. **Amortization into the base.** Once the prevalent-patient backlog is worked down, only new incident cases remain, so annual spend falls to a low steady state.

Year	2014	2015	2016	2017	2018	2019
HCV franchise revenue	~\$12.4B (ramp)	~\$19.1B (peak)	~\$14.8B	~\$9.1B	~\$3.7B	~\$2.9B

Table 5. Gilead HCV franchise revenue arc (Sovaldi + Harvoni + Epclusa) — the pulse shape: ~2-year ramp, 2015 peak, ~80% decay by 2018. **EST/RECALL** from Gilead 10-Ks;²² re-pull primary filings before client use.

Renewal-math takeaway: Hep C added a large *one-time* loading to 2014–2015 trend and then became a *tailwind* (a declining line item) in 2016–2019. A plan that baked the 2014 spike permanently into its base over-priced every renewal from 2016 forward. That error — treating a pulse as a permanent level — is the single most important lesson of the case.

Case 2 — GLP-1 obesity therapy: the persistent, growing wave

GLP-1s for obesity are the opposite of Hep C in every relevant dimension. They are **chronic** (indefinite use), the eligible population is enormous (~40%+ of US adults are obese), and weight regain on discontinuation pulls patients back — so there is no cure, no pool drawdown, and *no self-limiting decay*. IQVIA reports GLP-1 agonists contributed \$14B of growth in 2025 (\$9.6B in obesity indications) within a US net medicine market of \$606B (+10.6% YoY), with total net spending projected to grow +4.5–7.5%/yr through 2030.¹³ On the employer side, KFF EHBS 2025 shows GLP-1 weight-loss coverage at 16%/30%/43% of firms by size band, with the 5,000+ band jumping from 28% to 43% in a single year; among covering large firms, 66% report a “significant” Rx-spend impact and 59% saw usage higher than expected.²

What could bend the curve — but not turn it into a pulse this decade: oral versions (oral semaglutide, orforglipron) that raise volume but cut per-script price; the semaglutide patent cliff (~2031–2032) enabling biosimilar erosion; price competition (tirzepatide versus semaglutide, cash-pay programs); and coverage/eligibility tightening (prior authorization, BMI thresholds, step therapy).² These bend the *slope*; they do not shrink the eligible pool, so the wave persists.

The leading indicators, ranked by lead time

#	Indicator	Lead time	Source	Access
1	Drug pipeline / Phase III readouts	2–4 yrs	IQVIA; analyst pipeline; ClinicalTrials.gov	Free / paid
2	FDA approval calendar / PDUFA / novel-drug list	3–18 mos	FDA CDER novel-drugs list, PDUFA trackers	Free (FDA.gov)
3	PBM annual drug-trend forecasts (name next-yr drivers)	6–18 mos	Evernorth/Express Scripts, CVS Caremark, Optum Rx	Free (annual PDFs)
4	Prevalence × price × uptake (your own model)	1–3 yrs	CDC/NHANES + list/net price + penetration curves	Free + judgment
5	Cure-vs-chronic + biosimilar/LOE timeline	structural	FDA Orange/Purple Book; drug label	Free (FDA)
6	Early real-world script uptake (first 2–4 quarters)	3–9 mos	IQVIA/Symphony script data; own claims feed	Paid / internal

Table 6. Specialty-drug leading indicators, ranked by lead time and signal quality. The predictive punch comes from combining #1–2 (something big is coming) with #5 (pulse or wave?) and #3 (what PBMs say it costs next year), then quantifying with #4.

The “spike-then-level” model — an additive Rx-trend loading

The goal is a per-member-per-year Rx loading, $\Delta R_{x_{drug}}(t)$, added on top of baseline trend. The model is a six-step pipeline whose only branch is the cure-versus-chronic classification in step 1:

1. **Classify.** Cure (finite course, clears the condition) → a *pulse* with decay. Chronic (ongoing therapy) → a *persistent* plateau or continued growth. Signal: FDA label, mechanism, and whether treatment draws down the eligible pool.
2. **Eligible population.** $\text{Eligible}(t) = \text{Members} \times \text{Prevalence} \times \text{Clinical-eligibility fraction}$. For a *cure*, this is a stock that *depletes*: $\text{Eligible_cure}(t) = \text{Prevalent_backlog} - \sum \text{Treated}(<t) + \text{Incident_new}(t)$. For a *chronic* therapy it is replenishing/growing.
3. **Penetration S-curve.** Logistic uptake of the eligible pool, $\text{Penetration}(t) = \text{Ceiling} / (1 + e^{-(t-t_{\text{mid}})})$. For a chronic therapy the Ceiling is net of discontinuation/leakage — GLP-1 real-world one-year persistence is low, so model a “leaky bucket,” not a one-way fill.
4. **Net-price trajectory.** $\text{NetPrice}(t) = \text{List}(t) \times (1 - \text{Rebate}(t))$, with step-downs at each competitor launch and at loss of exclusivity / biosimilar entry. A cure is priced per-*course* (one-time); a chronic therapy per-*year* (recurring).
5. **Assemble the annual loading.**
 - Cure (pulse): $\Delta \text{Rx_cure}(t) = \text{Newly_treated}(t) \times \text{NetCoursePrice}(t)$ — ramps, peaks, then decays to an incident×price steady state as the backlog empties.
 - Chronic (wave): $\Delta \text{Rx_chronic}(t) = \text{Treated}(t) \times \text{Persistence}(t) \times \text{NetAnnualPrice}(t)$ — ramps to a high plateau, bending only on price erosion or eligibility tightening.
6. **Feed the renewal.** $\text{Projected_trend} = \text{Base_medical} + \text{Base_Rx} + \sum_{\text{drugs}} [\Delta \text{Rx_drug}(t+1) / \text{Total_expected_claims}]$, expressed as **named basis-point line items** so that a pulse can be *removed* as it decays — the Hep C lesson, encoded.

Sanity anchors from the case studies.

- Hep C calibrates the **pulse** shape: ramp 2014, peak 2015, ~80% decay by 2018 as the pool is cured and Mavyret crashes net price.
- GLP-1 calibrates the **persistent** shape: \$14B of growth in 2025 and still climbing; the bend will come from price and LOE, not from pool exhaustion.

Figure 4 places this overlay in the wider signal set: the specialty/GLP-1 pipeline is not only the longest-lead indicator but also one of the strongest, which is why Section 7 makes it the primary differentiator.

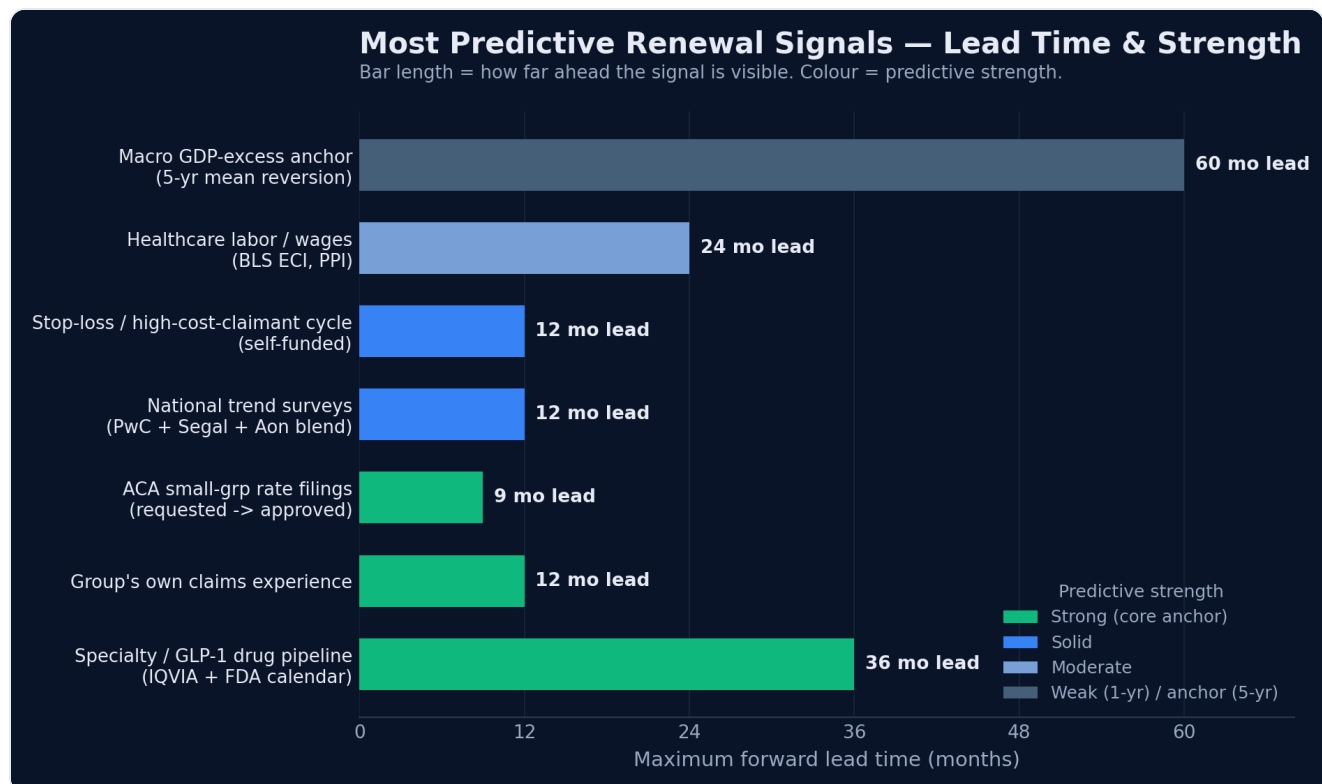


Figure 4. Leading indicators ranked by forward lead time and predictive strength. Bar length is how far ahead the signal is visible; color is predictive strength. The specialty/GLP-1 drug pipeline offers the longest genuine lead (up to ~36 months) with the cleanest public data, alongside the group's own claims as the strongest case-level signal.

7 Toward a better composite — standard + a differentiating factor

The winning design is not a new index. It is the standard trend plus one differentiating overlay.

No canonical *academic* leading-index for commercial renewals exists, and it would be a mistake to imply one to clients. The closest existing composites are the CMS NHE projections (the nearest to an official econometric composite),⁴ the Altarum Health Sector Economic Indicators (the nearest to a monthly leading index, blending health CPI/PPI, spending, and employment),⁷ and the Peterson-KFF tracker (a macro anchor).¹¹ The survey products (PwC, Milliman, Segal, KFF) are coincident, not leading. The productive move, therefore, is not to invent an index but to take a solid standard base and add a single differentiating overlay that carries genuine forward information.

The ranked added factors

#	Factor	Lead	Best source
1	Specialty / GLP-1 / cell-&-gene drug pipeline	12–36 mo	IQVIA, FDA novel-approval calendar, own PBM data
2	Stop-loss / reinsurance cycle + high-cost-claimant frequency	3–12 mo	Sun Life stop-loss report, Aon/Marsh; own stop-loss file
3	Healthcare labor / wages (unit-cost pressure)	12–24 mo	BLS ECI (healthcare) + CES earnings (free)
4	Hospital price + market concentration	6–18 mo	BLS hospital PPI + HHI (directional, not level)
5	Utilization (deferred/rebound, HCC freq, new CPT)	6–18 mo	Own claims
6	Macro (GDP-excess, CPI, wages)	weak 1-yr	Good 5-yr mean-reversion anchor only
7	Global / OECD / global reinsurance	—	Benchmarking only — the US <i>drives</i> global trend

Table 7. Added factors beyond standard trend, ranked by lead and data quality. The pharma pipeline leads the field; global indicators do not lead the US at all.

A note on hospital price (factor #4): the hospital PPI runs near +3.4%, far below the ~8%+ commercial trend, because it blends Medicare and Medicaid reimbursement. It is a useful *acceleration/direction* signal, never a level.³ **On the global question (factor #7):** foreign and OECD indicators mostly do *not* lead the US, because the US drives global medical trend rather than following it. Use them for multinational benchmarking, not for prediction. The one weak exception is that global pharma launch pricing can foreshadow US specialty list prices.

Three “standard + X” designs

A — PRIMARY

Base + pharma-pipeline overlay (1-yr & 5-yr)

Base = PwC/Segal/Milliman (1-yr) or CMS NHE (5-yr). X = additive basis points on the Rx component, driven by the book’s therapeutic-class exposure × IQVIA specialty/obesity net-price growth + FDA imminent-launch count + a GLP-1 penetration curve fit to group age/BMI (the Section 6 model). **The primary differentiator** — longest lead, cleanest data, and directly actionable at the case level.

B

Base + stop-loss / HCC loading (self-funded, 1-yr)

Base = actuarial claims trend. X = the group’s own stop-loss quotes and lasers + Sun Life/Aon high-cost-claimant frequency + a reinsurance hard/soft indicator → a large-claim loading and a leveraged-trend multiplier above the deductible. Best for self-funded cases, where own stop-loss quotes price 3–12 months ahead of the plan renewal.

C Base + macro / labor anchor with GDP mean-reversion (5-yr)

Base = CMS NHE / long-run. X = BLS healthcare ECI / wage acceleration + hospital PPI + local HHI + a GDP-excess constraint pulling the long run toward GDP + ~1.3 pp (capped at the historical ceiling). This is the Section 3 glide, operationalized as an overlay.

Implementation. Estimate the overlay weights by regressing historical renewals on lead-lagged features, using elastic-net regularization to handle the heavy collinearity among trend, wage, and price series — do *not* hand-set them. Overlay A (pharma) is the primary differentiator; B and C are situational additions for self-funded and long-horizon work respectively.

8 A recommended renewal-prediction architecture

Two engines, a five-year glide, a smoothed base rule, a specialty-drug overlay, and a proprietary calibration index.

Everything above converges on a single build. The architecture has two forecasting engines (selected by rating method, per Distinction 2), a shared five-year glide, the smoothed base rule established by the backtest, the specialty-drug overlay, and — the piece no competitor can copy — Benefitra's own book as the calibration index.

ENGINE 1 Small group — next-year, per-state (read the filings)

- Ingest the **CMS Rate Review PUFs** + ratereview.healthcare.gov (requested → approved), per state × carrier × year.^{9, 10}
- **Market-share-weight** by covered lives to a per-state small-group forecast; refresh each summer as filings publish.
- Use the **approved single-risk-pool change as the anchor**; **requested trend** as the early leading read; the **requested-versus-approved gap** only as a regime-gated secondary feature (interacted with a prior-approval flag).

ENGINE 2 Large group — next-year (blend × area × own claims)

- National trend blend (**PwC Group + Segal + Aon**, ~9% now) × per-state **area factors** (Milliman HCG if licensed, else BLS-metro CPI/PPI + CMS GPCI proxy).^{1, 14, 16}
- Move each account by its **own claims experience** — the dominant signal for any specific group.
- Calibrate against KFF/Mercer realized *and* Benefitra's own renewal book.^{2, 15}

5-YEAR The Getzen forward glide

- Getzen "GDP + X" glide: near-term ~7.5% → ~5–5.5% long-run; ~6.5–7.0% five-year national average.⁶
- State flex via CMS State Health Expenditure historical ratios (each state's spend-per-capita growth versus national as a multiplier).
- Always present with the **±1 pt/yr band** — never a bare point estimate.

BASE RULE Smoothed-recent-toward-mean (from the backtest)

- Where a pure time-series carry is needed (interpolation, gap-filling, sanity bounds), use **EWMA $\alpha = 0.5$** ($\approx 0.6 \times \text{last} + 0.4 \times \text{prior}$), nudged ~10–20% toward the ~6% long-run center.²⁴
- Never carry naive last-year forward, and never fall back to the pure long-run mean — both lost the backtest.

OVERLAY**The specialty-drug “spike-then-level” loading**

- Layer the Section 6 model on the Rx component as **named basis-point line items**, classified cure-vs-chronic so a pulse can be removed as it decays.^{13, 25}
- Add a **stop-loss / HCC loading** for self-funded cases (overlay B).
- Estimate overlay weights by elastic-net regression on lead-lagged features — not hand-set.

OWNED EDGE**Benefitra’s book as the calibration index**

Benefitra’s own **placed-case renewal book + PBM + stop-loss files** are the highest-value proprietary per-state signal in the entire stack. Structure them as an internal, continuously-updated index and calibrate both engines against it. This is the one input no competitor can replicate — and it converts every public signal above into a private advantage.

What is free/ingestible now versus what needs licensing or build**Free & ingestible now**

- CMS Rate Review PUFs + ratereview.healthcare.gov (small-group per-state)
- CMS NHE projections + State Health Expenditure Accounts
- Getzen/SOA long-run model (free Excel, annual)
- BLS CPI/PPI/ECI, CMS GPCI, FRED series
- KFF EHBS + Peterson-KFF tracker; Altarum HSEI
- FDA novel-approvals calendar; ClinicalTrials.gov
- PBM annual drug-trend PDFs (Evernorth, CVS, Optum)
- Benefitra’s own book / PBM / stop-loss files (owned)

Needs licensing or build

- PwC / Segal / Aon / WTW / Mercer trend surveys (re-pull & license before client use)
- Milliman Health Cost Guidelines (area factors)
- IQVIA specialty/obesity net-price series
- Sun Life / Aon stop-loss & HCC-frequency data
- Per-state prior-approval-vs-file-and-use classification table (build)
- The lead-lag elastic-net weighting pipeline (build)
- Citation-grade full KFF historical series (re-pull)

9 Does an indexed fully-insured renewal change what we recommend?

An impact test of the tool’s fully-insured renewal assumption. The short answer: it is a transparency-and-accuracy control, not a guidance-changer.

The architecture of Section 8 leans on a fully-insured (FI) renewal assumption as the benchmark every alternative funding path — PEO, Taft-Hartley, self-funded, self-funded-with-captive — is measured against. A natural governance question follows: if we swapped that FI assumption for an explicitly “indexed” value, would the tool’s five-year winning probabilities move, and — more importantly — would it ever change which funding path the tool recommends? We ran a controlled impact test to settle it. The result is decision-inert, and understanding *why* clarifies exactly what an FI-renewal control can and cannot do.

The mechanics that constrain the answer

Two facts about the engine bound the experiment before any simulation runs. First, the tool’s current FI-renewal default is **not a flat national number** — it is already a *per-state six-year filed average* (the STATE_RENEW_AVG table), so the fully-insured renewal is *already state-indexed*. Second, that input only reaches **community-rated FI** business — groups at or below each state’s community-rating threshold (50 lives in most states; 100 in CA, VT, and NY). **Experience-rated FI** (larger groups) ignores the input entirely and renews on the group’s own credibility-blended trend. The renewal is also expressible only as a single flat scalar per run — no per-year glide can be entered — so the strongest available “index” form is a single flat value.

Per-state filed FI-renewal defaults (STATE_RENEW_AVG, 6-yr filed average):

- NY 9.4% · MA 7.4% · TX 7.3% · FL 5.9% · AL 5.4% · WY 5.3% · CA 5.0% · national fallback 6.5%.
- Only **community-rated** FI ($\leq$ the state threshold: 50, or 100 in CA/VT/NY) reads this input. Experience-rated FI renews on its own trend.

Method

On a patched, enrolled-correct build of the engine we swept **35 segments** — {NY, CA, MA, FL, TX, AL, WY} × {30, 70, 150, 250, 500} lives — at N = 5,000 Monte-Carlo paths (seed 20260704, average age 42, health 1.0), holding the PEO renewal at 13% and the Taft-Hartley renewal at 3%. The **baseline** uses each state's per-state filed default; the **index** arm replaces it with a flat **6.75%** for every state — the five-year Getzen-glide average of Section 3, and the only form (a single flat scalar) the engine will accept. We then compared, segment by segment, each alternative path's probability of beating fully-insured over five years and, decisively, the tool's best-fit recommendation.

Result — decision-inert

0 / 35

segments change the best-fit recommendation. Every flip = false.

OWNED · index-impact sweep ²⁶

9 / 35

segments move at all — all community-rated (size 30 in every state, plus size 70 in CA & NY). The other 26 are $\Delta = 0$ by construction.

OWNED · index-impact sweep ²⁶

78–100%

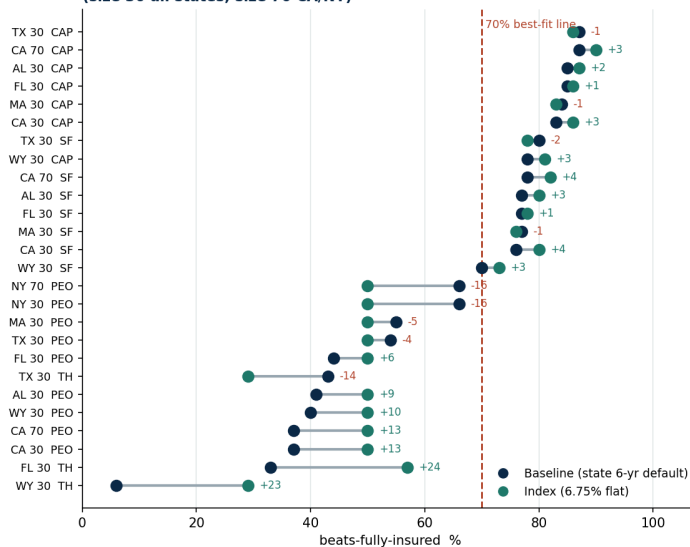
win probability of the recommended path under arms; **0** paths cross the 70% beats-fully-insured best-fit threshold.

OWNED · index-impact sweep ²⁶

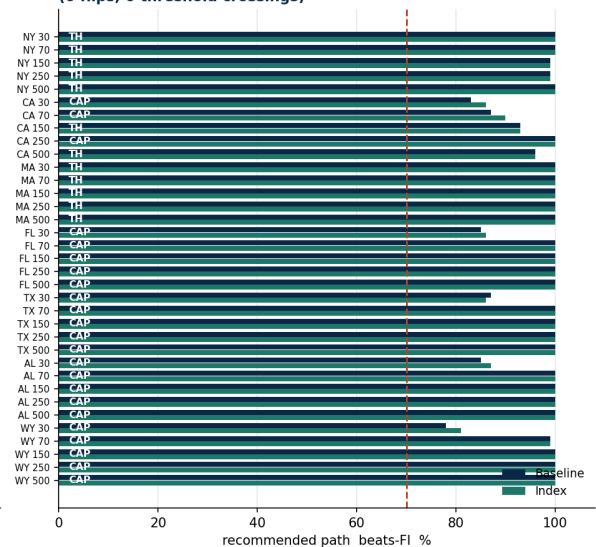
Only the community-rated segments move at all — every 101–500-life segment, and every 51–100-life segment outside CA and NY, is $\Delta = 0$ by construction because experience rating never reads the input. Among the nine that do move, the direction is *mixed, not uniformly down*: where the state's filed default sits *above* the 6.75% index (NY 9.4%, MA 7.4%, TX 7.3%), the index makes FI look *cheaper*, so alternatives beat it *less* (PEO -16 pts in NY); where the default sits *below* 6.75% (CA, FL, AL, WY), the index makes FI look *pricier*, so alternatives beat it *more* (Taft-Hartley +24 pts in FL-30, +23 pts in WY-30). The mean shift is about **+0.5 pt** with a **median of 0**. Crucially, none of these moves carries any path across the 70% best-fit line, so the recommendation is unchanged in all 35 cases.

FI-renewal index toggle: baseline (per-state) vs index 6.75% — 5-yr winning probabilities

Only community-rated segments move
(size 30 all states; size 70 CA/NY)



Best-fit pick holds in all 35 segments
(0 flips, 0 threshold crossings)



avg_age 42 · health 1.0 · seed 20260704 · N=5000 · PEO 13% / TH 3% held fixed. Experience-rated segments (all 101-500; 51-100 outside CA/NY) are insensitive by construction.

Figure 5. The FI-renewal index has no effect on guidance. *Left:* the nine community-rated segments that move at all, showing each alternative path's beats-FI% under the per-state filed default versus the flat 6.75% index, with the 70% best-fit line marked — note that no bar crosses it. *Right:* the recommended path's beats-FI% holding across all 35 segments under both arms (78–100%), with zero recommendation flips. Simulation: N = 5,000, seed 20260704, avg age 42; PEO 13% / TH 3% fixed. **OWNED**

The segments that moved

Segment	State filed FI default	Under 6.75% index	Recommended path (win% base → index)	Largest alternative Δ
NY · 30	9.4%	FI cheaper	Taft-Hartley (100 → 100)	PEO -16 pts
NY · 70	9.4%	FI cheaper	Taft-Hartley (100 → 100)	PEO -16 pts
MA · 30	7.4%	FI cheaper	Taft-Hartley (100 → 100)	PEO -5 pts
TX · 30	7.3%	FI cheaper	Captive (87 → 86)	TH -14 pts
FL · 30	5.9%	FI pricier	Captive (85 → 86)	TH +24 pts
AL · 30	5.4%	FI pricier	Captive (85 → 87)	PEO +9 pts
WY · 30	5.3%	FI pricier	Captive (78 → 81)	TH +23 pts
CA · 30	5.0%	FI pricier	Captive (83 → 86)	PEO +13 pts
CA · 70	5.0%	FI pricier	Captive (87 → 90)	PEO +13 pts

Table 8. The nine community-rated segments whose on-screen win-probabilities shift when the FI renewal is set to a flat 6.75% index. All are size 30, plus CA-70 and NY-70. Every recommendation is unchanged (flip = false); no path crosses the 70% best-fit threshold.

Verdict & implication

VERDICT

An indexed FI renewal is an accuracy control, not a guidance-changer

Switching the fully-insured renewal to an indexed value moves the on-screen small-group win-probabilities but **never changes which funding path the tool recommends** — 0 flips across 35 segments, and it touches community-rated groups only. Treat an FI-renewal index as a *transparency-and-accuracy* setting, not a lever on advice.

The test also carries a design warning. A single *flat national* index (the 6.75% used here) would be **less accurate** than the existing per-state default: it overstates the alternatives' odds in high-trend states (NY, MA, TX) and understates them in low-cost states (CA, FL, AL, WY) — exactly the mixed-direction pattern the sweep exposes. Any indexed setting should therefore stay **per-state-aware** rather than collapse to one national figure. That is precisely what the current STATE_RENEW_AVG table already is — which **validates the paper's recommendation** (Sections 2 and 8) to keep the small-group FI renewal anchored to per-state filed data. The finding is thus a confirmation, not a course-correction: the engine already indexes the one place indexing helps, and does not spend the change where it would not.

10 Limitations & data provenance

The honesty notes, consolidated. Read this before any figure leaves the building.

SOURCED

Verified live during research — figure and URL confirmed.

ESTIMATE

Analyst blend or modeled figure — not a single published number. Labeled as such throughout.

RE-PULL

Proprietary vendor figure (PwC / Segal / Aon / WTW / Milliman / Sun Life / IQVIA) referenced from research but **blocked live** — must be re-pulled and licensed before any client-facing use.

OWNED

Benefitira proprietary data (placed-case book, PBM, stop-loss) — the calibration layer.

Consolidated limitations

- The ~9% next-year anchor and all survey trends are **RE-PULL**. PwC, Segal, Aon, WTW, Mercer, Milliman, IQVIA, and Sun Life figures were referenced from research but were not all fetched live this session; several vendor sites blocked automated access. Re-pull and license the current-year figures before any client-facing use.

- **The five-year 6.5–7.0% blend is an ESTIMATE.** Its inputs (CMS, CBO, Getzen, Altarum, BEA parameters) are sourced, but the specific yearly glide points and the state spread are modeled, not published. Present with the ± 1 pt/yr band; Getzen's own stated accuracy is $\sim \pm 1\%$ /yr, compounding to $\sim \pm 5.1\%$ over five years.
- **The backtest rests on 18 out-of-sample points, and its interior series years are EST.** Recent and cumulative KFF values were confirmed live; interior years are training-knowledge estimates reconciled to those cumulatives. The top nine rules are statistically bunched — the robust claim is "any smoothing beats naive-last and pure-mean," not that EWMA 0.5 is uniquely best. The series is also an all-firms family-premium *proxy*, not a literal small-group-rated book.
- **The specialty-drug case figures are largely EST/RECALL.** Sovaldi's price/date, the Gilead franchise arc quote, and the GLP-1/KFF/IQVIA employer figures are sourced; but the CMS +12.2% (2014), the Express Scripts +742%, the ASPE $\sim \$12.3$ B, and the year-by-year Gilead franchise revenues are recalled industry figures. Re-pull the primary government and PBM sources before client use. Figure 3 is an explicitly **stylized schematic** — the shapes are calibrated to the documented arcs, but the exact yearly values are estimates.
- **The requested-vs-approved gap requires a per-state regime table that has not yet been pulled.** Without the prior-approval-vs-file-and-use classification, the gap feature cannot be correctly gated.
- **Large group has no filed data — a hard limit, not a gap to close.** The requested-vs-approved machinery of Section 4 simply does not apply; large group relies on the survey blend, area factors, and own claims.
- **All per-state five-year detail is flexed from national figures.** CMS/CBO/Getzen forecasts are national-only; state numbers are national trend multiplied by historical state ratios, not independent state forecasts.

Every vendor and recalled figure in this paper is a placeholder for a licensed re-pull. The method is the deliverable; the numbers are illustrative until sourced for the client.

11 References

Grouped by type. All URLs are drawn from the underlying research briefs.

Government & official (free, sourced)

1. **PwC Health Research Institute**, "Medical Cost Trend: Behind the Numbers 2027" (Group 9.0%, Individual 8.5%). [pwc.com/.../behind-the-numbers.html](https://www.pwc.com/.../behind-the-numbers.html) **RE-PULL**
2. **KFF**, Employer Health Benefits Survey 2025. [kff.org/.../2025-employer-health-benefits-survey](https://www.kff.org/.../2025-employer-health-benefits-survey)
3. **US Bureau of Labor Statistics**, Consumer Price Index — medical care components (Jul 2025 – Jul 2026); hospital PPI FRED series PCU622110622110. [bls.gov/cpi](https://www.bls.gov/cpi)
4. **CMS Office of the Actuary**, National Health Expenditure Projections 2025–2034 (forecast summary). [cms.gov/.../nhe-projections](https://www.cms.gov/.../nhe-projections)
5. **Congressional Budget Office**, Budget & Economic Outlook and Long-Term Budget Outlook (excess cost growth). [cbo.gov](https://www.cbo.gov)
6. **Getzen / Society of Actuaries**, Long-Run Medical Cost Trend Model, 2025 update. [soa.org/programs/health-care-cost-trends](https://www.soa.org/programs/health-care-cost-trends); 2025 update at [soa.org/.../2025-getzen-model-update](https://www.soa.org/.../2025-getzen-model-update)
7. **Altarum**, Health Sector Economic Indicators (HSEI). [altarum.org/health-sector-economic-indicators](https://www.altarum.org/health-sector-economic-indicators)
8. **Bureau of Economic Analysis**, PCE health price index (FRED DHLCRG3M086SBEA). fred.stlouisfed.org/...
9. **CMS**, Single-Risk-Pool Rate Review Public Use Files (2014–2026). [cms.gov/marketplace/resources/data/rate-review-data](https://www.cms.gov/marketplace/resources/data/rate-review-data)
10. **CMS / CCIIO**, RateReview.HealthCare.gov (requested + approved justifications). ratereview.healthcare.gov
11. **US Food & Drug Administration**, CDER novel drug approvals list & PDUFA calendar; Orange/Purple Book (LOE dates). [fda.gov/.../novel-drug-approvals](https://www.fda.gov/.../novel-drug-approvals)

Tracker, survey & actuarial

11. **Peterson-KFF Health System Tracker**, premium briefs (2025, 2026) and Rx-spending trends. [healthsystemtracker.org](https://www.healthsystemtracker.org)
12. **Georgetown CHIR**, 2026 rate-filing analysis (178 filings, 28 states + DC). [chir.georgetown.edu](https://www.chir.georgetown.edu)
13. **IQVIA Institute**, Use of Medicines in the US 2025. [iqvia.com/.../use-of-medicines-in-the-us](https://www.iqvia.com/.../use-of-medicines-in-the-us)
14. **Segal**, Health Plan Cost Trend Survey. [segalco.com](https://www.segalco.com) **RE-PULL**
15. **Aon; WTW; Mercer**, employer/trend surveys (Mercer National Survey of Employer-Sponsored Health Plans). **RE-PULL**
16. **Milliman**, Medical Index (MMI) & Health Cost Guidelines (area factors). [milliman.com/.../milliman-medical-index](https://www.milliman.com/.../milliman-medical-index) **RE-PULL**
17. **Sun Life** stop-loss report; **Aon/Marsh** high-cost-claimant benchmarking. **RE-PULL**

Specialty-drug case sources

- 18. **Wikipedia**, "Sofosbuvir" (Sovaldi price/date, 46% discount, 2020 net course cost). en.wikipedia.org/wiki/Sofosbuvir
- 19. **Wikipedia**, "Gilead Sciences" (revenue arc; "number of suitable patients decreased" quote). en.wikipedia.org/wiki/Gilead_Sciences
- 20. **Express Scripts**, 2014 Drug Trend Report (Hep C +742% YoY). **EST/RECALL — RE-PULL**
- 21. **ASPE / HHS**, "Observations on Trends in Prescription Drug Spending," 2016 (~\$12.3B Hep C, 2014). **EST/RECALL — RE-PULL**
- 22. **Gilead Sciences**, Forms 10-K, 2014–2019 (HCV franchise year-by-year revenue). **EST/RECALL — RE-PULL**

Benefitra internal research (this program)

- 24. **Benefitra Actuarial Intelligence**, "Backtest — best predictor of next-year small-group renewal" (Part 1), 2026. Artifacts: analyze.py, build_series.py, results.json, renewal_predictor_backtest.png. **OWNED**
- 25. **Benefitra Actuarial Intelligence**, "Specialty Medications as Leading Indicators of Renewal Increases" (Part 2), 2026. **OWNED**
- 26. **Benefitra Actuarial Intelligence**, "Indexed Fully-Insured Renewal — Engine Impact Test" (35-segment sweep), 2026. Artifacts: run.js, results.json, index-impact-chart.png. **OWNED**