The Case for Mandating Digitisation of Health Records: Evidence from a Prototype Risk Score

Agastya Bhandari

Abstract

India lacks a standardised, individual-level health risk metric equivalent to the CIBIL credit score, leaving insurers to price risk largely on self-declaration and occasional point-in-time screening. This paper constructs a prototype health risk score for the Indian market — as a diagnostic exercise rather than a commercial product — to identify the data gap and quantify the cost of the data gap. We compare a Full Clinical Model incorporating laboratory biomarkers against a Restricted Predictor Model without them; the clinical model achieves a PR-AUC 33% higher, with six of the ten most predictive variables originating from laboratory results. Yet in practice fewer than 5% of policyholders undergo screening at underwriting, and existing medical records remain largely un-digitised and fragmented. We argue that this data deficit sustains a self-reinforcing "Market-Failure Feedback Loop": pervasive information asymmetry prevents refined risk pricing, forces an implicit cross-subsidy from healthy to less-healthy policyholders, suppresses low-risk participation, and surfaces as claim denials that deepen a public trust deficit and entrench under-penetration. Drawing on the precedent of credit bureaus in Indian lending, the paper contends that closing this gap requires a government mandate for standardised health-data reporting. It further examines the central risk such a mandate creates — exclusion of high-risk individuals — and outlines protections (risk pooling, policy design, constrained use, and universal-coverage backstops) intended to ensure that fuller disclosure leads to fair pricing rather than denial of coverage.

Executive Summary

India currently operates without a standardized, individual-level health risk metric equivalent to the CIBIL credit score. Introducing a centralized health score holds immense utility for predictive underwriting, corporate wellness programs, and proactive consumer health engagement. This paper attempts to construct a prototype health risk score for the Indian market. Developed as a diagnostic exercise rather than a commercial product, the model empirically maps the systemic data gaps currently constraining the industry.

Our findings reveal a critical structural bottleneck for Indian insurers, who face a severe deficit of laboratory data, despite its high predictive value. Specifically, the Full Clinical Model (FCM)—which integrates laboratory biomarkers—achieved a PR-AUC 33% higher than the Restricted Predictor Model (RPM). This sharp performance gap is driven by the fact that six of the top ten predictive variables originate from laboratory results. In the real world, though, fewer than 5% of policyholders undergo medical screening during underwriting, equally few seek regular preventive health check-ups, and existing medical records remain largely un-digitised and fragmented. This baseline reality stands in stark contrast to the global benchmarks examined later in this paper —such as Medicare Risk Adjustment Factor and Dacadoo Health Score— which explicitly leverage continuous, comprehensive lab data to accurately quantify underwriting risk.

The primary obstacle to building an India Health Score (IHS) is data availability. Unlike the credit market, India has no national mandate for reporting medical information to a centralized registry. Furthermore, health data is inherently more complex than lending data; instead of a borrower's straightforward payment updates, medical records are voluminous, unstandardized, and plagued by low fill rates that degrade data integrity and predictive validity.

Indian health underwriting relies overwhelmingly on self-declaration and occasional, point-in-time medical screenings. This profound information asymmetry, compounded by a scarcity of initial and longitudinal policyholder data, prevents the deployment of advanced predictive pricing models. The resulting lack of refined risk pricing creates an implicit cross-subsidy from healthy to less-healthy individuals, which simultaneously suppresses low-risk market penetration and triggers an adverse selection loop. Because underwriting remains anchored to good-faith disclosures, structural friction frequently surfaces as claim denials during settlement. Each rejection deepens the public trust deficit, directly crippling the organic pull of health insurance despite a glaring structural need reflected in severe under-penetration despite high medical inflation. This is the Market-Failure Feedback Loop in health underwriting. Ultimately, cultivating trust is the critical variable required to convert health insurance from a product that must be sold into one people actively seek, mirroring the transition that turned mutual funds into a mass-market “pull” product in India.

To bridge the current trust deficit and realize affordable universal healthcare, policy interventions must establish a centralized data stack capable of powering advanced health underwriting technologies and drive the transition of the Indian health insurance industry towards dynamic, predictive models. Equitable risk assessment for fair pricing to all policyholders requires high-quality longitudinal data that is both granular and standardized. Preserving data integrity requires comprehensive reporting rather than selective or optional participation, highlighting the absolute need for government mandates integrated with robust technological architecture.

The risk with mandatory sharing of health records is denial of insurance to high-risk customers. The interest of the applicant can be protected via a) Pool funding b) Policy Design and/or c) Constrained use of shared health records. An alternate approach is for the government to extend Base Universal Health Coverage to all (extended PM-JAY scheme), that solves the problem of denial to any applicant; applicant can choose to enhance the policy at his/her own expense. Experience from the credit market offers a measure of reassurance: individual credit scores did not, overall, lead lenders to exclude low-scoring borrowers, but to price them appropriately. In a market with robust regulatory oversight and healthy competition, risk-based pricing rather than outright exclusion is the more probable equilibrium for health as well. One critical difference between credit risk and health risk prevents us from relying on that equilibrium alone: unlike a loan, where downside is bounded by the amount advanced, a catastrophic health claim carries a potentially unbounded cost over the life of the policy — an exposure created by guaranteed renewability and only partly offset by cohort repricing. This means that a small tail of genuinely high-risk customers may remain uninsurable at any commercially viable price. It is precisely this tail — not the broad middle of the risk distribution — that the mitigations above are designed to protect. Risk-based pricing handles the many; the pool, policy-design and universal-coverage backstops handle the few for whom pricing alone would still mean exclusion.

The Ayushman Bharat Digital Mission (ABDM) is India's national digital health initiative that seeks to create an integrated, interoperable digital health infrastructure, allowing citizens to securely access and share their digital health records (like prescriptions, lab reports). The Government and the IRDAI have established foundational platforms such as the Insurance Information Bureau (IIB), alongside work-in-progress initiatives like the National Health Claims Exchange (NHCX) and the Public Information Registry (PIR). The power of the platforms is dependent on the quality of data – granularity and reliability. Currently, the absence of a strict regulatory mandate compelling healthcare providers to share data in a standardized format severely constrains adoption and deployment. Driving meaningful collaboration across fragmented stakeholders will therefore require a firm, government mandate, mirroring the historic intervention that originally established credit bureaus in the financial sector.

Introduction

India lacks an individual health risk metric equivalent to the CIBIL credit score. Medical data remains trapped in paper files and siloed hospital systems, forcing insurers to price risk based entirely on good faith and self-reporting. Consequently, corporate wellness programs lack a structured, principled methodology to prioritize urgent preventive care, while individuals remain without a simple, portable metric to track their health trajectories over time.

Lenders faced a similar structural problem around 2 decades ago: a lender could not easily tell which applicants were likely to default, because relevant financial history including repayment records was scattered across banks. India's financial sector solved this asymmetry through the CIBIL score, a single number between 300 and 900 that compresses years of loan and repayment behaviour into a figure any lender can act on within seconds. Health has no equivalent number calibrated for Indian patients. The consequences of miscalculating risk are far more profound than in lending due to the regulatory architecture of health insurance, which mandates guaranteed renewability and restricts premium repricing to the cohort level rather than the individual level.

Globally, several organizations and jurisdictions have introduced unified health scores that condense an individual's overall health status and risk profile into a single metric, mirroring the utility of the CIBIL score in the financial sector. These scoring systems are increasingly deployed across insurance underwriting, corporate wellness initiatives, and consumer health engagement platforms.

We analyse three global case studies illustrating the application of unified health scores:
- Medicare’s Risk Adjustment (United States): A nationwide, regulatory-driven framework utilized for population health risk stratification.
- Dacadoo Health Score: A digital health scoring system launched by Dacadoo; partnership with Narayana Health for One Health Score; early-stage implementation that has yet to achieve broad market scale.
- Binah.ai: An innovative technology platform that leverages contactless video processing to extract vital statistics, offering alternative mechanisms for health data acquisition

Figure 1: Summary of Health Score Case Studies

	Key Use	Key Data Variables	Data Source
Medicare’s Risk Adjustment Factor (USA)	Single risk score used in U.S. Medicare Advantage to predict each enrollee’s relative health-care costs and adjust plan payments accordingly	Demographics Diseases (mapped to ICD-10)	Hospital inpatient and outpatient, and physician claims

Dacadoo Health Score	Used by insurers as a way to engage customers, assess risk and incentivise healthy behaviour. Also used in corporate wellness and healthcare provider programs.	Demographic	
		Clinical Lab data	Health records
		Family history	
		Lifestyle	Self-declaration Wearable devices
		Psychological	Self-declaration
Binah.ai (Israel)	Health data platform that produces a single Wellness Score	Deployed mainly in life/health insurance and wellness programs. Insurers mainly use the score for quick, non-invasive pre-underwriting checks	Uses a short front-camera scan (computer-vision and signal-processing techniques) to extract vitals such as heart rate, heart-rate variability (HRV), respiratory rate, oxygen saturation (SpO_2), and a cuffless blood-pressure proxy, and other vitals

1. Medicare's Risk Adjustment Factor (USA)

The Risk Adjustment Factor (RAF) is the single risk score used in U.S. Medicare Advantage to predict each enrolee's relative health-care costs and adjust plan payments accordingly. It's a unified number designed for payment fairness across populations with different morbidity burdens

The inputs required for RAF are sourced from CMS enrolment files and provider-documented encounters/claims (only conditions that are monitored, evaluated, assessed, or treated in the year count).

In Medicare Advantage, the Risk Adjustment Factor (RAF) expresses an enrolee's expected cost of care relative to the program average. It is a decimal value calibrated around 1.0. Any figures above 1.0 signal higher-than-average expected spend (for example, a member with diabetes and complications might be ~2.0), while figures below 1.0 indicate lower risk (e.g., a very healthy member might be ~0.6). CMS applies each enrolee's RAF to the benchmark/base payment to determine plan reimbursement and thereby align plan payments with clinical risk.

CMS calculates the RAF prospectively by combining two primary input classes to estimate an enrolee's relative cost for the coming year

- Demographic baseline: age and gender, disability/original-entitlement status, residence segment (community vs. institutional), new-enrolee models (when prior-year data are incomplete).
- Diagnoses mapped to HCCs: ICD-10 diagnosis codes from the base year are grouped into Hierarchical Condition Categories (HCCs); hierarchy rules ensure more severe conditions supersede related, less-severe codes to prevent double counting.
- Specified interactions/segments: defined disease combinations and model segments add incremental weight where clinically relevant.

Illustrative Example (End-to-End)

Consider a concrete example similar to our running scenario. Suppose an enrollee is a 76-year-old female in the community, not dual-eligible, with a history of Type 2 diabetes, chronic congestive heart failure, and atrial fibrillation (and no other HCC conditions). Under V28, her risk calculation would look like:

- Demographic factor (Female 75–79, non-dual, community): 0.1949.
- Mapped HCCs: Diabetes (HCC #37 in V28), CHF (HCC #226), Atrial fibrillation (HCC #238), for example.
- HCC coefficients: Diabetes = 0.166; CHF = 0.360; Atrial fibrillation = 0.299.
- Sum of HCCs: 0.166 + 0.360 + 0.299 = 0.825.
- Add demographic: 0.825 + 0.1949 = 1.0199 (raw subtotal).
- Disease interaction: Diabetes & CHF interaction adds 0.112.
- New subtotal: 1.0199 + 0.112 = 1.1319 (raw risk score).
- HCC count factor: She has 3 HCCs; count factors start at 5, so 0.000 added.
- Raw risk score remains: 1.1319.
- Divide by normalisation 1.045: 1.1319 ÷ 1.045 = 1.083 (normalised risk).

- Apply 5.9% coding adjustment: 1.083 × 0.941 = 1.019 (final RAF).
- After rounding, RAF = 1.019.

This means her risk is about 2% above average. If the county's base payment for MA is, say, $10,400, the plan would get roughly $10,400 × 1.019 ≈ $10,597 for her

2. Dacadoo Health Score (Switzerland)

Dacadoo is a digital health platform provider that offers real time health scores ranging from 0 to 1000 that dynamically represents overall health. Primarily this score is used by insurers to engage customers, assess risk and incentivize healthy behaviour. It is also used in corporate wellness and healthcare provider programs.

In 2024, Dacadoo partnered with Narayana Health to launch the One Health Score (OHS) for Indian consumers. By consolidating diverse health data into an easy-to-understand number, the OHS aims to guide individuals on a personalized path to better health and preventative care.

Dacadoo's health score combines data from three main data domains

- Body (biometric data): age, sex, height, weight (BMI), waist circumference, clinical lab values (cholesterol, blood glucose, etc.), as well as family medical history
- Lifestyle: physical activity (steps, heart rate, workouts), nutrition (diet quality, alcohol intake), sleep quality, smoking status, and substance use. Much of this data can be captured via wearable devices and smartphone sensors integrated with the app
- Mind (Psychological): self-reported mental well-being, stress, mood, social engagement, mindfulness practices

These data inputs come from manual entry into the app, or through wearables, smartphone sensors and linked health records

The score engine is built on 9+ years of research and draws from 100+ epidemiological models plus clinical study data. It uses established medical risk equations—the kind that link BMI, blood pressure, and cholesterol to outcomes like disease and mortality—and calibrates these models with machine learning from 400 million person-years of anonymized clinical and lifestyle data. This allows the system to continuously refine risk predictions as more users generate data. The machine-learning layer keeps tuning the model as new data arrives, re-weighting factors when evidence changes—for example, if research shows sleep quality has a bigger impact than previously thought, the sleep sub-score is adjusted.

The OHS is not a static once a year generated score, but rather dynamically updates with new data. By bridging fragmented health data into one score, OHS seeks to "simplify healthcare" for its users which help them lead healthier lives.

3. Binah.ai (Israel)

   Binah.ai is a software-only health data platform that uses a short front-camera scan to extract vitals and produce a single Wellness Score (1–10). It is deployed mainly in life/health insurance and wellness programs. Insurers mainly use the score for quick, non-invasive pre-underwriting checks which gives an instant health-risk assessment from a selfie rather than a medical exam.

   The Wellness Score is a single unified number displayed alongside the underlying vitals—heart rate, heart-rate variability (HRV), respiratory rate, peripheral oxygen saturation (SpO_2), and a cuffless blood-pressure estimate—so the drivers of the score are visible.

   Vitals extraction (rPPG): Remote photoplethysmography with computer-vision and signal-processing techniques transforms subtle facial color changes into a clean pulse waveform. From this, the system estimates the heart rate, heart-rate variability (HRV), respiratory rate, oxygen saturation (SpO_2), and a cuffless blood-pressure proxy, and other vitals.

   Feature mapping and composite scoring: The measured vitals are converted into risk features (for example, HRV balance, oxygenation stability, BP range, breathing pattern). These are then compared with clinically informed reference ranges and are combined using weighting and normalization and converted into a single 1–10 Wellness Score.

Indian Health Score (IHS)

India lacks a standardized individual health risk metric equivalent to the CIBIL credit score. This gap stems from a severe deficit of real-world health data available to both insurers and policyholders. This paper attempts to construct a prototype health risk score for the Indian market. Serving as a diagnostic framework rather than a commercial application, the model highlights the systemic data gaps currently constraining the industry.

This paper develops two predictive models: the Full Clinical Model (FCM), which incorporates demographics, anthropometric variables, biomarker variables from laboratory reports and pre-existing disease profiles, and the Restricted Predictor Model (RPM), which utilizes the same dataset but excludes laboratory reports. The primary objective is to evaluate whether FCM can meaningfully differentiate high-risk from low-risk patients based on hospital admission rates over the subsequent six months given access to differential data (laboratory reports).

The data was sourced from a health ecosystem partner and de-identified before analysis. It was accessed for research under the custodian's existing consent and the provisions of the DPDP Act, and no re-identification was attempted. Due to data access constraints, the scope of this exercise remains limited. Consequently, these findings should not be interpreted as a validated clinical or actuarial instrument. The absolute figures are illustrative and are not intended as population-representative or actuarial estimates.

Algorithm:

We evaluated various model constructs – Logistic Regression (LR), Generalised Linear Models (GLM), Random Forest (RF) and Extreme Gradient Boosting (XG Boost)

- Logistic Regression: A specific, entry-level case of a GLM. It assumes a strict linear relationship between your input health metrics (like age or blood pressure) and the log-odds of a health event
- GLM (Generalized Linear Model): The broader statistical framework enclosing logistic regression. It lets you model non-normal health outcomes, like healthcare costs (via Gamma distribution) or hospital visit counts (via Poisson distribution).
- Random Forest: An ensemble machine learning model that builds dozens of independent, parallel decision trees. It averages their outputs to catch complex, non-linear patient risk factors without overfitting
- XG Boost: A highly optimized sequential tree-boosting system. Each new decision tree focuses entirely on correcting the mistakes of the previous trees, making it the most accurate but hardest to interpret without external tools

Figure 2: Comparison of Models : Design and Tradeoffs

Metric / Criteria	Logistic Regression	GLM (Generalized Linear Model)	Random Forest	XG Boost
Primary Output Type	Binary probability (0 to 1)	Flexible (Counts, Cost, Binary)	Probability or Continuous	Probability or Continuous
Handling Missing Data	Fails (Requires manual imputation)	Fails (Requires manual imputation)	Robust (Handles missing clinical values)	Excellent (Learns optimal default paths)
Predictive Accuracy	Moderate (Misses complex interactions)	Moderate (Dependent on link function)	High (Captures non-linear traits)	Very High (Top tier accuracy)
Explainability / Trust	Excellent (Clear clinical risk odds)	Excellent (Clear mathematical weights)	Moderate (Requires feature importance)	Low but SHAP (Shapley Additive explanation) or LIME (Local Interpretable Model-agnostic Explanations) can help explain the model

Regulatory Compliance	Very Easy (FDA / HIPAA friendly)	Very Easy (Highly auditable)	Harder (Requires explainability layers)	Hardest (Requires strict validation)

Because this paper serves a diagnostic function rather than proposing a regulated commercial product, predictive accuracy was prioritized over strict regulatory explainability. Consequently, we selected XG Boost for its high predictive performance and paired it with SHAP values to interpret key variables.

Parameters:

We consulted health experts to reflect industry and India-specific nuances such as

- Period
    - Recent data points are more relevant (Observation Window)
    - Hospitalisation over next 12 months (Performance Window) is a better outcome to model instead of Health Claim because it is highly predictive, mathematically stable, and insulates the model from volatile fluctuations in medical billing inflation. Due to data limitations, the analysis was constrained to hospitalization outcomes within a six-month window.
    - Factor in Execution Window (say 1 month) between Observation Window and Performance Window
- Data variables
    - We excluded direct, one-to-one variable mappings
    - Algorithmically, income significantly improves predictions for long-term health outcomes. Statistically, lower-income populations exhibit higher chronic disease prevalence, reduced treatment adherence driven by out-of-pocket expenses, and elevated rates of avoidable hospitalizations. Income directly correlates with a patient's access to preventative nutrition, high-quality housing, safe working environments, and chronic stress levels. However, income was excluded from the final model due to data quality constraints: specifically, a low fill rate, poor data integrity due to misreporting, and high-income volatility characteristic of India's largely self-employed workforce.

Our algorithm was trained on a population of more than 300,000 with more than 80 variables and an average 3.26% claim rate (over next 6-month period).

These variables were filtered to have Information Value (IV) and high Fill Rate (a constraint specific to our data source).

Results

Our trained model was then tested on a population of around 80,000. The results of the same are tabulated below

Note: This exercise is diagnostic: it isolates the marginal predictive contribution of laboratory data by comparing FCM and RPM on the same cohort. The absolute figures are illustrative and are not intended as population-representative or actuarial estimates

Figure 3: Model Performance by Risk Decile (FCM vs. RPM): Event Rates

Risk Decile	Population	Full Clinical Model (FCM)			Restricted Predictor Model (RPM)		
		Events[(1)]	Event Rate	% of Overall Events	Events[(1)]	Event Rate	% of Overall Events
1	8,000	618	7.68%	23.54%	524	6.5%	20.0%
2	8,000	382	4.74%	14.55%	357	4.4%	13.6%
3	8,000	301	3.74%	11.47%	303	3.8%	11.5%
4	8,000	279	3.46%	10.63%	258	3.2%	9.8%
5	8,000	236	2.93%	8.99%	216	2.7%	8.2%
6	8,000	207	2.57%	7.89%	233	2.9%	8.9%
7	8,000	193	2.40%	7.35%	214	2.7%	8.2%
8	8,000	172	2.14%	6.55%	195	2.4%	7.4%
9	8,000	132	1.64%	5.03%	179	2.2%	6.8%
10	8,000	105	1.30%	4.00%	146	1.8%	5.6%
Top 3 deciles				49.6%			45.1%
Bottom 5 deciles				30.8%			36.8%

(1) Event = Hospitalisation over the next 6 months

Precision Recall – Area Under Curve (PR-AUC) was 2.1x the baseline in FCM (0.0689 vs 0.0326), which is considered good. This was 33% better than PR-AUC in RPM (1.6x).

In FCM, targeting the top decile, 10 percent of members, captures 23.54% of all future claimants at 2.35x lift in FCM. The top three high risk deciles, 30% of the population, capture close to half of all claimants in FCM. Monotonic decay across deciles confirms the score ranks risk consistently end to end.

On a comparative basis, RPM reported lower events in the top 2 high risk deciles and materially higher events in the bottom 2 high risk deciles. While monotonic decay generally holds across the table, the same fails between Decile 5 and 6. Consequently, RPM would end up mis-pricing policyholders

Note that health insurers operate with low underwriting profit (as % of Revenue) and hence high sensitivity to appropriate pricing

Below are reported financials of Star Health, India's largest standalone health insurer

Figure 4: Financial Results and Claim Cost - Star Health (India's largest standalone health insurer)

Rs bn	FY26	
Insurance Revenue	180	
Claims	123	Event Rate = 6-7%
Claims Ratio	**68.2%**	Frequency x Severity
Combined Insurance Service Ratio	98.4%	
Underwriting Profit	1.6%	
Profit After Tax	**5.1%**	Includes Investment Income
Return on Equity	**13.1%**	

Source: Company Disclosures

<u>Top variables</u>

Figure 5: Top Predictive Variables — Full Clinical Model (ranked by |SHAP|)

Parameter Ranking	Feature	Category / Source	|SHAP|[(1)]	Comment
1	Age	Demographic	0.2102	
2	BMI	Underwriting	0.0929	
3	Serum Albumin	Laboratory Report	0.0750	Vital biomarker measured via a standard blood draw to evaluate liver function, kidney health, and systemic nutritional status
4	Weight	Underwriting	0.0728	
5	Package Name	Policy Details	0.0630	
6	HbA1c	Laboratory Report	0.0629	Average blood sugar (glucose) levels over the past 2 to 3 months
7	Gamma-glutamyl transferase	Laboratory Report	0.0626	Vital enzyme found primarily in the liver

8	Blood Urea	Laboratory Report	0.0565	Marker for Kidney Health
9	Thyroid-Stimulating Hormone	Laboratory Report	0.0544	Critical master hormone produced by pituitary gland to control body's metabolism
10	Absolute Neutrophil Count	Laboratory Report	0.0537	Neutrophils are your body's primary line of defense against infections

(1) |SHAP| is a metric used in machine learning to rank how much impact different features have on a model's final predictions. It measures the overall importance of a variable by averaging the absolute strength of its influence across every single data point in a dataset

Figure 6: Top Predictive Variables — Restricted Predictor Model

Parameter Ranking	Feature	Category
1	Age	Demographic
2	Weight	Anthropometric variables
3	Number of Pre-Existing Diseases	Pre-Existing Diseases
4	Location	Geography
5	BMI	Anthropometric variables
6	Height	Anthropometric variables
7	Gender	Demographic

Age dominates, followed by a metabolic and organ function cluster: HbA1c, serum albumin and markers for liver, kidney and pituitary gland such as Gamma-Glutamyl Transferase, Blood Urea and Thyroid-Stimulating Hormone respectively in FCM.

Figure 7: Share of Predictive Contribution by Data Source — Full Clinical Model (ranked by |SHAP|)

Source	Total SHAP	Share
Laboratory Report	2.0326	80%
Demographic	0.2499	10%
Underwriting	0.1944	8%
Policy Detail	0.0630	2%

This underscores the critical importance of laboratory data

- Six of the top ten features in the model were derived from laboratory results.
- Disclosures regarding specific pre-existing diseases (PED) – Diabetes, Hypertension, Musculoskeletal, Renal, Surgical and Thyroid had negligible uplift given availability of laboratory data.

Our Learnings

The effort of the paper was to map the systemic data gaps currently constraining the industry. Comparison of the results across the 2 models – FCM and RPM, clearly underscore the importance of laboratory data.

While 6/10 top variables are from laboratory results, insurers have limited access to laboratory results as less than 5%[1] of the policyholders undergo medical test at the time of underwriting (policy issuance), equally few seek regular preventive health check-ups, and existing medical records remain largely un-digitised and fragmented.

While the previously discussed frameworks—Dacadoo and Medicare Risk Adjustment—rely on direct laboratory reports to calculate a health score, Binah.ai adopts an alternative methodology. Utilizing computer vision and signal processing, its remote photoplethysmography technology extracts vital signs from subtle variations in facial coloration. Collectively, however, all three frameworks underscore the critical role of real-time physiological data in enhancing underwriting precision.

On the other hand, Indian underwriting is overwhelmingly self-declaration and occasionally, a point-in-time medical test. This manifests as severe challenges for the health insurance industry

- Data deficits and structural fragmentation severely constrain underwriting sophistication in the Indian health insurance market. Information asymmetry, along with the scarcity of both initial and longitudinal policyholder data, prevents the deployment of advanced predictive pricing models

- The lack of refined pricing creates an implicit cross-subsidy where healthy individuals inadvertently subsidize less healthy lives. As a result, higher-risk individuals are onboarded below fair market rates, while healthier individuals choose to opt out due to inflated premiums. This dynamic not only suppresses overall health insurance penetration but also triggers severe adverse selection, ultimately driving sharper premium increases for the remaining insured cohort.

- Because underwriting leans on self-declaration and good faith, disputes frequently surface at claim settlement; Denials at claim time, arising from "non-disclosure" disputes after the policyholder has paid premiums during the hour of need is the worst possible experience for the policyholder. If everything is disclosed upfront through mandatory sharing, the non-disclosure allegation largely evaporates: there's nothing hidden to contest. So mandatory sharing can reduce the denials that destroy trust (claim-stage repudiation) even as it enables more sophistication at issuance.

- India’s lengthy waiting and moratorium periods create significant friction for consumers. This restrictive environment directly suppresses demand, keeping health insurance a pushed product rather than a pull product.

  <u>Disclosures Rely Entirely on Good Faith</u>

Because medical examinations are typically waived for younger age groups to lower friction during sign-ups, the entire policy issuance framework operates on the "Principle of Utmost Good Faith." The 5-year moratorium acts as a balanced legal compromise: it grants insurers a reasonable window to detect intentional, hidden medical fraud while ensuring that consumers who cross the 5-year mark achieve complete policy incontestability.

The Absence of Centralized, Interconnected Electronic Health Records (EHR)

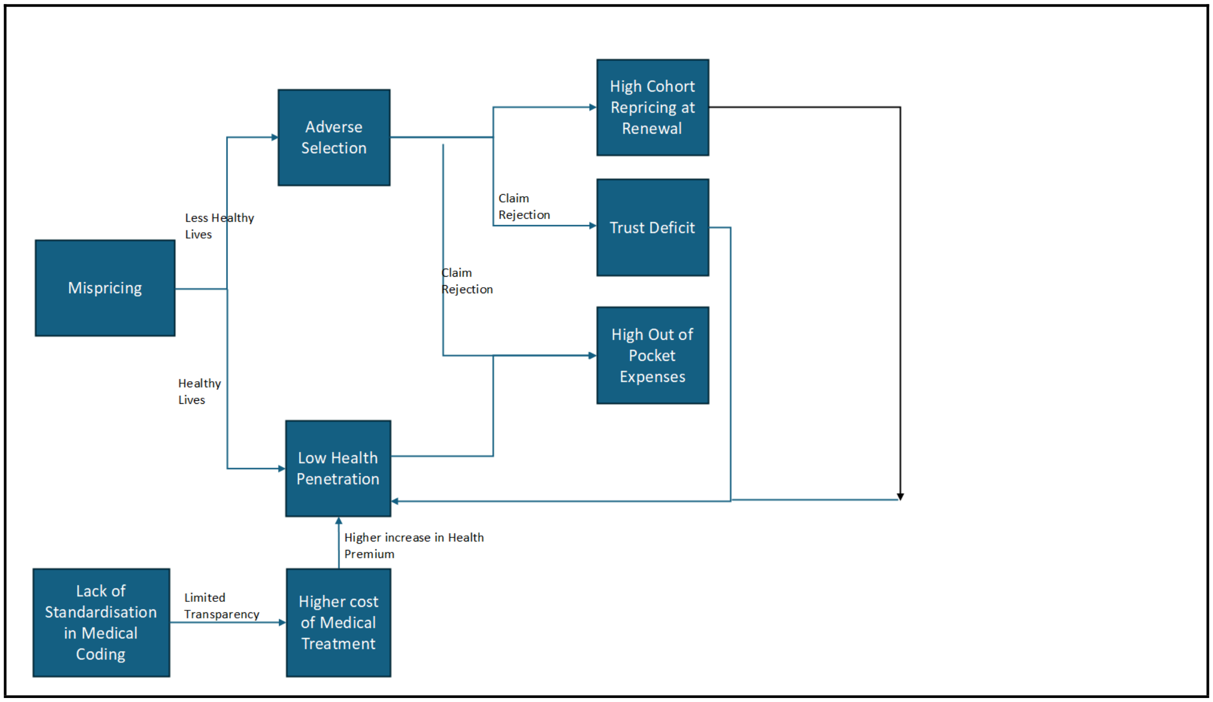
In nations like the UK or Singapore, an insurer can instantly cross-verify an applicant's complete historical medical timeline using a centralized national health registry. In India, healthcare data remains highly fragmented across thousands of independent private clinics and hospitals. Because insurers cannot easily audit an applicant's medical history during initial underwriting, they rely on a prolonged 5-year moratorium window to observe and catch fraudulent patterns.

Figure 8: Comparison of PED Waiting Period and Moratorium across geographies

Geography	Max Pre-Existing Diseases Waiting Period	Moratorium / Incontestability Concept
United States[1]	0 Days (Banned under ACA)	Not applicable; background health history cannot void a plan.
United Kingdom[2]	No fixed wait; dependent on moratorium tier.	2-Year Rolling window of no symptoms/treatment required to earn coverage.
Australia[3]	12 Months	Covered by community-rating laws; no retroactive cancellations.
Singapore[4]	Can be Permanent for private plans.	Occasional 2 to 5-year rolling windows offered by select private vendors.
India[5]	3 Years (Capped by IRDAI)	5 Years (Cannot reject for non-disclosure after this)

(1) Section 2704 of the Public Health Service Act
(2) Association of British Insurers (ABI)
(3) Private Health Insurance Act 2007
(4) Singapore's Ministry of Health (MOH), Life Insurance Association (LIA) Singapore for Integrated Shield Plans.
(5) IRDAI Master Circular on Health Insurance Business (29 May 2024), IRDAI (Insurance Products) Regulations, 2024, effective 1 April 2024.

Figure 9: A Market-Failure Feedback Loop in Health Underwriting



Building Data Stack Foundation for Longitudinal Care

Bridging the trust deficit to realise affordable healthcare for all will require a centralised data stack that will enable leveraging technology for superior health underwriting.

Data required for a fair policy underwriting by the health insurer requires

Longitudinal Data	For data to be relevant and useful, it will require continuous digital reporting of data instead of selective data at the choice of the policyholder/individual
Granularity of Data	Leveraging predictive AI requires standardized, granular, and continuous electronic health records (EHR). The greater the capture of clinical history, diagnostic data, and lifestyle indicators, the greater is the ability to leverage AI/ML. The machine-learning layer keeps tuning the model as new data arrives, re-weighting factors when evidence changes. It therefore requires granularity of data capture as patterns evolve or new patterns emerge
Standardisation of Data	A critical hurdle to building these AI training models is the current lack of a unified language across hospitals and between hospitals and insurers. Systemic data capture requires the mandatory adoption of standard

	classification metrics and uniform medical coding, such as ICD-11 (International Classification of Diseases) codes across all healthcare providers.
Household level	In India, around 60%[1] of the policies are family floater policies. Therefore, the underwriting at the time of policy issuance ideally requires household and not just individual underwriting.

Challenge of Disclosure

Greater disclosure is generally beneficial, but it carries a risk for the applicant: fuller disclosure could lead insurers to deny coverage to high-risk customers altogether. Interest of the applicant can be protected in a few ways as we discuss below.

The below options are not exclusive, and regulator/policy makers can choose to adopt a combination of the below

a) Pool funding

- As a case study, it is interesting to revisit the regulatory history wrt Third Party Insurance for Commercial Vehicles in the 2006-2016 period
    - o Before 2007, general insurers hesitated to cover commercial vehicles due to exceptionally high claim frequencies and massive financial risks. Since Third-party motor liability insurance is legally mandatory in India, the pool (as discussed below) was set up to prevent insurers from turning away commercial vehicle owners.
    - o The third-party pool concept is like Indian Motor Third Party Insurance Pool (IMTPIP) in Commercial Vehicles in India (from Apr 2007 till Mar 2012), wherein all registered general insurance companies collectively pooled their third-party commercial vehicle premiums and shared the claims liabilities proportional to their total market share.
    - o This was replaced by Indian Motor Third Party Declined Risk Insurance Pool (from Apr 2012 till Mar 2016), wherein only for high-risk, declined vehicles (like heavy trucks), the pool absorbed the risk, and all the general insurers contribute to payouts based on their total market share
- Adapting the above framework to Health Insurance to ensure that an individual is not denied health coverage
    - o Insurers can't deny coverage to high-risk customers but can tag them as pool customer
    - o For these pool customers, part premium is funded by government or by the health insurance industry (proportionate to their market share). The part premium is the excess to cover the actual Loss Ratio above the maximum premium payable by non-pool customer. The customer still pays the maximum premium payable by non-pool customer
    - o In case the government funds the part premium (instead of the health insurance industry), this will be in a way extension to the PM-JAY scheme

- o The underwriting models by the health insurers will be auditable to ensure that they are not tagging excessive policies as pool policies, funded by government or third-party pool
- o As in the case of Motor Third Party Pool, as the healthcare ecosystem matures, the above pool can be modified or phased out.

b) Policy Design

- Insurers can offer policy with co-pay, deductible or narrow network to these high-risk customers
- These policy design features can help reduce premium by 10-30% individually and if applied cumulatively, can results in larger savings in premium.

Figure 10: Affordable Premium Under Co-Pay and Deductible Options (Illustrative)

	Detail	**Annual Premium (Rs)**	**Saving**
Premium	1 year Sum Insured = Rs 10 Lakh	23,561	
Voluntary Co-Payment	10-30% co-pay	(2,509 – 7,396)	11-31%
Deductible	5-10% of Sum Insured i.e. 50,000 – 1 Lakh	(3,962 - 6,603)	17-28%

Source: www.policybazaar.com
Policy: Niva Bupa, Reassure 3.0 Elite Value, Family of 4 with highest age of 49 years

c) Constrained Use of Health Records at time of Issuance

To ensure that customers are not denied coverage, shared access to longitudinal health records can be kept (i) optional or (ii) avoided completely at the time of policy issuance. Access to health records can be made mandatory at the time of renewal for cohort repricing, care management and claim-time verification.

In case the framework adopts the optional route, withholding consent can itself be an underwriting-relevant signal, allowing insurers to price the absence of disclosure conservatively. Non-disclosure will therefore cease to be a costless way to conceal risk, which substantially weakens the incentive to opt out. In that case though, to ensure non-denial of coverage, this framework will need to be complemented by one of the above options a) or b).

While this limits the benefits of access to longitudinal health records, it is a healthy compromise while solving for more than one objective.

An alternate approach is for the government to extend Base Universal Health Coverage to all (extended PM-JAY scheme), that solves the problem of denial to any applicant.

As a case study, the Singapore model has some similarity with the mixed healthcare model that may be appropriate for India

- Singapore operates a multi-layered "S + 3Ms" financing framework that guarantees comprehensive medical access to all citizens and permanent residents (PRs) while enforcing strict individual cost-sharing to prevent systemic waste. This ensures that nobody is denied necessary medical attention due to inability to pay.
- Subsidies and basic insurance are strictly optimized for shared, basic public hospital wards (Class B2 and C). If a patient prefers private hospitals or air-conditioned single rooms (Class A), the public coverage drops significantly, requiring private "Integrated Shield Plans" to bridge the financial gap.
- There is no 100% free care. Patients always pay a small portion of their bills through deductibles or co-insurance to discourage over-consumption of healthcare resources

S – Government Subsidies	The government provides direct tax-funded subsidies of up to 80% on medical bills at acute public hospitals and polyclinics
M1 – MediShield Life	A mandatory, national, catastrophic health insurance plan that covers all citizens and PRs for life, regardless of pre-existing conditions. It is specifically designed to absorb large hospital bills and expensive outpatient regimes (like chemotherapy or dialysis)
M2 – MediSave	A mandatory, individual medical savings account. Working individuals and their employers must legally contribute 7% to 10.5% of their monthly salary into this fund. These funds are locked and can only be withdrawn to pay for personal or immediate family hospitalizations, surgeries, and specific chronic outpatient care
M3 – MediFund	A state-backed medical endowment fund acting as an absolute bottom safety net. If a citizen completely exhausts their subsidies, MediSave, and MediShield Life, MediFund steps in to pay the remaining balance so that no one goes untreated

An adaption of the above can be implemented by extending the government's flagship PM-JAY health insurance scheme to the entire population

- This will meet the Universal Health Coverage commitment
- The coverage under the current PM-JAY scheme is for Rs 0.5mm p.a. for a lower income household. The government can choose to offer a limited coverage (i.e. co-pay/deductible) for the middle income and above household to manage the fiscal burden on government
- Households can choose to enhance the coverage by adding rider plan, wherein longitudinal health records will play an important role.

Recommendation: Government Mandate for Central Health Repository

Ayushman Bharat Digital Mission (ABDM) aims to establish a unified, interoperable digital health infrastructure through which citizens can securely access and share their medical records, including prescriptions and lab reports. The digital backbone (Ayushman Bharat Digital Mission) can serve as the "digital highway" connecting the various healthcare stakeholders, including patients, doctors, and hospitals, for seamless, continuous care across the country and superior underwriting outcomes for the health insurance industry.

Preserving data integrity requires comprehensive reporting rather than selective or optional participation. The foundational principle therefore requires mandatory reporting at the provider layer with consent-gated access at the insurer layer. To transition Indian health insurance from static, limited pricing variables into dynamic, predictive risk models, the government must mandate data capture standards and data sharing protocols. Standardisation of medical protocols and uniform medical coding can help bring about higher transparency, standardization of care and care pathways and improve affordability. This will likely result in lower rejection and lower surprise of out-of-pocket expense burden on the policyholder/patient.

The risk with mandatory sharing of health records is denial of insurance to high-risk customers. Despite robust regulatory oversight and a competitive market, a small tail of genuinely high-risk customers may remain uninsurable at any commercially viable price. It is this tail — not the broad middle of the risk distribution — that the mitigations such as a) Pool funding b) Policy Design and/or c) Constrained use of shared health records besides an alternate approach for the government to extend Base Universal Health Coverage to all (extended PM-JAY scheme) are designed to protect.

The system architecture must incorporate robust safeguards for data privacy and consent-based sharing. The Ayushman Bharat Digital Mission (ABDM) framework directly aligns with the Digital Personal Data Protection (DPDP) Act. Under this model, patients link their medical records to a unique Health ID (ABHA), and data consumers must secure explicit, patient-notified approval via a designated Consent Manager before retrieving any linked records. This will help reconcile the mandate for reporting in a standardised format at the provider layer while keeping access consent-gated at the insurer layer

Government and IRDAI have existing platforms such as insurance Information Bureau (IIB) and work-in-progress platforms such as NHCX (National Health Claim Exchange) and PIR (Public Information Registry). The value of these platforms depends directly on the quality of the underlying data — both the granularity captured and its reliability. The absence of a regulatory mandate compelling healthcare providers to share data in a standardised format has, however, significantly constrained adoption and roll-out. Collaboration across stakeholders will therefore require a firm government mandate, mirroring the historic intervention that established credit bureaus in the financial sector.

References

1. The Policybazaar National Health Insurance Report
2. Digital Personal Data Protection: https://www.meity.gov.in/static/uploads/2025/11/53450e6e5dc0bfa85ebd78686cadad39.pdf
3. NITI Aayog Report : Health Insurance for India's Missing Middle
4. Ajay Mahal, Vivek Panwar, Arun Tiwari, Rahul Reddy and Sumit Kane : The 'Missing Middle': How to Provide 350 Million Indians with Health Coverage?
5. E&Y Parthenon: FICCI, True accountable care, October 2025
6. MercerMarsh Benefits: Health Trends 2026, Balancing Cost Control with Benefit Optimization
7. IRDAI Master Circular on Health Insurance Business (29 May 2024)
8. IRDAI (Insurance Products) Regulations, 2024, effective 1 April 2024.
9. Star Health, Reported Financial Disclosure