EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

EMPLOYEE WELLBEING INTERVENTION AND THE ECONOMIC COST OF PRODUCTIVITY LOSS: FROM RETROSPECTIVE COST MEASUREMENT TO PREDICTIVE WELLBEING ANALYTICS

HRM VARIABLE: WELLBEING PROGRAMME DESIGN → ECONOMIC OUTCOME: COST OF LOST OUTPUT

Dr. N Subbukrishna Sastry[1], Dr. Manjula Mallya M[2]

[1]Professor, School of Management, CMR University, Bangalore, Karnataka, India.
ORCID ID: 0009-0009-0460-7057,
[2]Associate Professor & Head, Dept. of Economics, Government First Grade College for Women Balmatta Mangalore, Karnataka, India. ORCID ID: 0009-0005-8812-6912,

ABSTRACT

Employers in the United States and comparable developed economies lose an order of magnitude more output to impaired and absent working time than they spend on the wellbeing programmes intended to recover it, yet the return to those programmes remains one of the least securely established quantities in applied human resource economics. This article argues that the difficulty is not primarily a difficulty of intervention design but of measurement regime. The prevailing regime is retrospective: it prices output loss after the loss has occurred, through claims run-out, absence registers and annual survey instruments whose reporting lag typically runs from nine to twenty-four months. A systematic review of 74 studies, of which 31 supplied extractable quantitative parameters, establishes three economic consequences of that regime. First, attribution failure: the reported return to wellbeing expenditure falls monotonically as identification strengthens, from 3.27:1 in observational meta-analysis to 1.50:1 in quasi-experimental employer studies and to no statistically detectable effect in the two large randomised trials, whose confidence intervals exclude the great majority of prior estimates. Second, latency: retrospective measurement can price a loss but cannot avert it, so the recoverable share of the loss decays with detection lag. Third, denominator error: presenteeism, the larger component of impaired output, is the component measured least precisely, so programme design optimises against the smaller and better-measured claims margin. Predictive wellbeing analytics is then assessed as a forward-looking alternative. The review finds that its economic value derives not from prediction as such but from targeting, which is the single feature common to the interventions that have survived rigorous evaluation. That value is bounded by the base rate of the risk detected, by the efficacy of the interventions the score triggers, and by a privacy and trust cost that is economic rather than merely ethical. The article concludes that neither regime is sufficient alone, and sets out a policy architecture in which prediction performs the targeting and embedded prospective randomisation performs the evaluation, supported by organisational rather than individual risk features, a restored federal incentive standard, and disclosure of measurement provenance alongside claimed returns. A comparative extension to India shows which parts of the argument are general and which are artefacts of American benefit financing: attribution failure, measurement latency and the base-rate bound on targeting all transfer, while the private return to investment is lower where sickness cost is partly socialised through the Employees' State Insurance scheme and the great majority of workers lie outside any measured perimeter. Because the claims apparatus that drew American programme design toward the wrong margin is largely absent, India's organised sector is positioned to adopt predictive targeting with embedded evaluation without first replicating the retrospective regime.

KEYWORDS — *Employee Wellbeing; Productivity Loss; Presenteeism; Cost Of Lost Output; Predictive Analytics; Return On Investment; Measurement Regime; Workplace Wellness; Human Resource Economics; Programme Evaluation*

I. INTRODUCTION

The economic case for employer investment in employee wellbeing rests on an accounting identity that is rarely disputed and a causal claim that is rarely established. The identity is that ill health, distress and disengagement reduce the output an employer obtains from a given wage bill. The causal claim is that expenditure on wellbeing programmes reduces that loss by more than the expenditure costs. The identity is supported by measurement of considerable scale. The Integrated Benefits Institute, drawing on benchmarking data covering approximately 147 million American workers and 66,000 employers, places the annual cost of poor health to United States employers at roughly US$575 billion and 1.5 billion lost workdays, equivalent to sixty-one cents of lost output for every dollar the same employers spend on health benefits [1]. At the global level Gallup values the productivity forgone to low engagement at approximately US$10 trillion, or nine per cent of world gross domestic product [2]. The causal claim, by contrast, has weakened steadily as the evidence base has been subjected to stronger identification.

This article treats that divergence as its object of study. It argues that the instability of the estimated return to wellbeing intervention is a property of the measurement regime within which the return is calculated, rather than of the interventions themselves. Under the prevailing regime, which this article calls retrospective cost measurement, the cost of lost output is

reconstructed after the fact from medical and pharmacy claims, disability and leave registers, absence records and periodic self-report instruments. The regime is coherent as accounting. It is weak as decision support, for three reasons developed at length below: it cannot separate programme effect from participant selection; it observes losses only once they have been incurred and are therefore no longer avertible; and it measures its largest cost component, impaired performance at work, with far less precision than its smaller and more legible claims component.

Predictive wellbeing analytics is advanced by vendors, and increasingly adopted by large employers, as the forward-looking alternative. It substitutes continuously observed operational signals — workload and scheduling data, collaboration and communication load, leave and assistance-programme utilisation patterns, and where consent permits, physiological data from wearable devices — for the periodic backward-looking instruments of the retrospective regime, and applies statistical learning to generate individual or team-level risk scores in advance of the loss. The claim attached to this substitution is that it converts a sunk cost into an avoidable one. This article assesses that claim against the assembled evidence and finds it partly sustained, materially qualified, and dependent on design choices that most current implementations do not make.

1.1 Research questions

The study is organised around four questions. First, what is the composition of the cost of lost output in developed-economy employment, and which components does the retrospective regime actually observe? Second, how does the estimated return to wellbeing intervention vary with the strength of the identification strategy used to estimate it, and what does that variation imply about the informational content of reported returns? Third, what does predictive wellbeing analytics change in the economics of intervention, and what bounds that change? Fourth, what measurement and policy architecture would allow employers, insurers and regulators to price the return to wellbeing intervention credibly enough to allocate capital against it?

1.2 Contribution

The article makes four contributions. It reframes an intervention-design debate as a measurement-regime debate, and shows that the reframing explains the pattern of findings in the literature more economically than the alternatives. It identifies targeting, rather than prediction, as the active economic ingredient in the interventions that have survived randomised evaluation, and derives from that identification a bound on the value of predictive analytics that depends on the base rate of the condition being predicted. It treats the privacy and trust cost of continuous wellbeing surveillance as an economic cost that enters the return calculation directly, through its effect on disclosure and therefore on data quality, rather than as an ethical externality lying outside it. And it specifies a hybrid architecture — predictive targeting with embedded prospective randomised evaluation — that inherits the lead time of the predictive regime and the attribution properties of the experimental one.

II. REVIEW OF LITERATURE

2.1 The cost of lost output as an economic category

Lost output is not a single quantity but a family of quantities with different observability. Illness-related absence removes labour input entirely and is recorded in payroll and leave systems. Disability and workers' compensation losses are recorded in insurance systems. Presenteeism — attendance at work with impaired capacity — leaves no administrative trace at all and must be elicited by instrument, a distinction developed at length by Johns [27] and given its first large monetised treatment for United States employers by Goetzel and colleagues [28]. Turnover attributable to wellbeing failure destroys firm-specific human capital and surfaces only as replacement cost. The Integrated Benefits Institute decomposition assigns 978 million days annually to illness-related absence and 540 million days to impaired performance, with an average of 10.3 lost workdays per employee per year distributed across sick leave, short- and long-term disability, workers' compensation and family and medical leave [1].

Placing this category correctly within labour economics matters for what follows. Lost output is not a welfare loss confined to the firm; it is a reduction in the effective labour input available to the economy, and it interacts with skill composition, labour market tightness and the pace of technological displacement. Sastry and Mallya [3] develop this connection in their treatment of strategic human resource management as an instrument of economic resilience, arguing that skill diversification and labour market adaptability determine how much of a productivity shock a workforce absorbs rather than transmits. Their framing is directly applicable here: a workforce whose effective input is already depressed by unmeasured impairment has less absorptive capacity when a technological or demand shock arrives, so the cost of lost output should be read as a resilience variable and not only as a benefits-line expense. The present article extends that reading by asking what the firm would have to measure in order to act on it.

A second structural point concerns the unit at which output loss accumulates. Aggregate per-employee averages conceal the fact that impairment is produced by particular configurations of work. Sastry [4], examining the effect of departmental structure on task achievement, shows that achievement varies systematically with how responsibility, reporting and task boundaries are arranged across units rather than with the attributes of individuals holding posts within them. The implication for cost measurement is significant and under-recognised: if the productive unit at which loss is generated is the department or team, then measurement systems built at the level of the individual claimant will attribute to persons what is properly attributable to structures, and interventions targeted at persons will be aimed at the wrong margin.

2.2 The retrospective measurement apparatus

The instruments of the prevailing regime are well documented. Direct cost is captured from medical and pharmacy claims, disability payments and workers' compensation reserves. Indirect cost is captured from absence registers and, for presenteeism, from self-report scales, principally the World Health Organization Health and Work Performance Questionnaire [5], the Work Limitations Questionnaire and the Stanford Presenteeism Scale. Health risk appraisals and biometric screening supply the risk-factor distribution against which programmes are designed. The Business Group on Health reports that mental health now features in the wellbeing strategy of every surveyed large employer, with financial health in ninety-two per cent, and that employers increasingly rely on wellbeing dashboards to evaluate vendor performance [6].

Each of these instruments is periodic and backward-looking. Claims data are complete only after run-out, typically three to six months beyond the incurral period, and are analysed annually. Health risk appraisals are administered once a year, usually tied to an enrolment window. Engagement surveys are annual or, at best, quarterly. The consequence is a reporting lag between the onset of impairment and its appearance in management information that runs from roughly nine months for absence and claims reporting to eighteen months or more where an annual survey is combined with claims run-out.

The periodicity is not incidental; it is inherited from the appraisal and reporting calendar of the human resource function itself. Sastry and Selvarasu [7], in their study of performance appraisal and employee engagement, show that engagement responds to the appraisal process and to the feedback it carries, and that the appraisal cycle therefore functions as the effective sampling frequency at which the organisation learns about its own workforce state. Read against the present argument, their finding identifies the origin of the measurement lag: wellbeing information is collected on the appraisal calendar because that is the calendar on which the function is administered, not because that is the frequency at which impairment develops. The mismatch between the two frequencies is precisely the quantity that predictive analytics proposes to close.

2.3 The evidential gradient: from observational return to randomised null

The most consequential development in this literature over the past fifteen years has been the systematic weakening of the estimated return as identification has strengthened. Baicker, Cutler and Song [8], meta-analysing largely observational employer studies, reported medical cost savings of US$3.27 and absenteeism savings of US$2.73 for each dollar spent. The RAND workplace wellness study [9], using quasi-experimental employer data, reported an overall return of approximately 1.50:1, but decomposed it sharply: the disease management component, taken up by thirteen per cent of employees, returned 3.80:1 and generated some eighty-seven per cent of total savings, largely through a reduction of close to thirty per cent in hospital admissions, while the lifestyle management component, taken up by eighty-seven per cent, returned 0.50:1 and reduced absenteeism by slightly more than one hour per employee-year.

Two large randomised trials then removed most of what remained. Song and Baicker [10] randomised 160 worksites of a national retailer, comparing 4,037 employees at twenty treatment sites with 28,937 at 140 control sites, and found that of eighty pre-specified outcomes at eighteen months only two differed significantly: self-reported regular exercise, higher by 8.3 percentage points, and active weight management, higher by 13.6 percentage points. Ten clinical markers, thirty-eight measures of spending and utilisation and three employment outcomes — absenteeism, tenure and job performance — showed no significant differences. Extended follow-up to three years reported persistence of the behavioural differences without the emergence of effects on clinical measures, health care spending or employment outcomes [11]. The Illinois Workplace Wellness Study [12] randomised nearly five thousand university employees and found persistent increases in screening participation but no significant effects on medical spending, health behaviours, productivity or self-reported health at twenty-four months; critically, its confidence intervals were tight enough to rule out roughly eighty-four per cent of previously published estimates of effects on medical spending and absenteeism.

The Illinois study also supplied the mechanism that reconciles the two bodies of evidence. In the year preceding the intervention, employees who went on to participate already had lower medical expenditure and healthier behaviours than those who did not [12]. Observational designs that compare participants with non-participants therefore attribute to the programme a cost differential that existed before the programme did. Selection, not intervention, generates a substantial part of the return reported under the retrospective regime.

Why individual-level programmes should perform so weakly is illuminated by work on the organisational production of strain. Sastry [13], in a study of leadership and workforce burnout, locates the proximate causes of burnout in leadership practice, workload allocation and the quality of managerial support rather than in the health behaviours of the individuals who exhibit it, and argues that revitalising leadership capability is the operative lever. If that account is correct, then a programme that offers screening, coaching and lifestyle modules to individuals is intervening downstream of the causal process, which is a sufficient explanation for its failure to move spending, absence or performance in randomised settings. The evidential null is not evidence that wellbeing does not matter economically; it is evidence that a particular class of intervention does not reach the mechanism.

2.4 What survived: targeting as the active ingredient

Two findings in the assembled evidence run against the general null, and they share a feature. The RAND disease management component returned 3.80:1 while its lifestyle counterpart returned 0.50:1 [9]; the difference between them is not intensity of expenditure but selection of recipient, since disease management is delivered to an identified population with a diagnosed and costly condition. At the population level, Chisholm and colleagues [14] estimated that scaling up treatment for depression and anxiety across thirty-six countries would cost US$147 billion and yield a five per cent improvement in labour

force participation and productivity valued at US$399 billion, with benefit-to-cost ratios of 2.3 to 3.0 on economic returns alone and 3.3 to 5.7 when health returns are included, against a global productivity loss from these conditions of the order of US$1 trillion annually. Here too the intervention is clinically indicated and delivered to identified cases.

A third observation points the same way. Deloitte's modelling of the United Kingdom employer cost of poor mental health, which placed the annual total at GBP 53 to 56 billion and the average return at GBP 5 for each GBP 1 invested, found that proactive designs outperform reactive ones and that presenteeism is the largest single cost component [31]. The proactive-reactive margin is a statement about timing, and timing is precisely what the predictive regime alters.

The generalisation available from these two results is that returns concentrate where intervention is targeted on identified need, and dissipate where it is offered uniformly to a population whose average risk is low. This is a statement about the economics of allocation rather than about the efficacy of any particular clinical or behavioural technology, and it is the premise on which the assessment of predictive analytics in Section V depends. Sastry [4] provides the organisational counterpart to the clinical case: where task achievement is shown to depend on structural configuration, the identified unit of need is a department with a particular arrangement of responsibility and reporting, and targeting means directing intervention at that unit rather than distributing it evenly across headcount.

2.5 Job design and job resources as the locus of intervention

If targeting is the active ingredient, the question becomes what the intervention should consist of once a target is identified. The demand-control tradition following Karasek [25] and the job demands-resources tradition consolidated by Bakker and Demerouti [26] hold that strain arises from the imbalance between demands placed on a worker and the resources and discretion available to meet them, which makes job resources and autonomy the employer-controllable levers. Two studies by Sastry and Ravish supply direct evidence on both levers in an employment setting.

The first [15] examines the empowerment of employees through delegation and the provision of resources, and finds that the two operate jointly: delegated responsibility unaccompanied by resource provision raises demand without raising the capacity to meet it, whereas the combination raises engagement and effectiveness. This is an important qualification for wellbeing programme design, because a substantial class of workplace interventions — autonomy initiatives, flexible working, self-managed scheduling — delegates discretion without adjusting the resource envelope, and would on this evidence be expected to increase rather than reduce strain. The second [16] examines delegated authority and resource provision as determinants of decision-making and employee initiative, and finds that where authority is genuinely devolved and matched with resources, employees exercise initiative rather than escalating decisions upward. The economic significance for the present argument is that the decision latitude which the demand-control tradition treats as protective is shown to be an operational design variable that management sets, and therefore a variable a measurement system could in principle observe continuously and act upon before impairment develops.

Taken together with Sastry [13], these results locate the recoverable portion of the cost of lost output in the design of work rather than in the health behaviour of workers. That location has a direct measurement implication. Workload allocation, span of control, schedule volatility, decision latitude and resource adequacy are all observable in operational systems that the employer already owns, and observable continuously rather than annually. They are also, unlike biometric and claims data, not health information, which materially alters the governance position discussed in Section 2.7.

2.6 Predictive wellbeing analytics: capability and claim

The predictive proposition is that risk of impairment can be estimated in advance from signals already generated by the conduct of work, and that intervention triggered on that estimate averts loss that retrospective measurement could only have priced. The signal set typically includes scheduling and overtime patterns, communication volume and out-of-hours activity, leave-taking sequences, assistance-programme and teletherapy utilisation, internal mobility and, where consented, wearable-derived sleep and activity data. Statistical learning methods are then used to produce risk scores at individual, team or unit level.

Sastry and Mallya [17], in their study of artificial intelligence analytics applied to hiring, provide the closest available treatment of this class of system within the human resource function and one that is instructive precisely because it is favourable in its overall assessment. They document substantial improvements in the efficiency of screening and in the quality of candidate matching, and identify the mechanism as the capacity of analytics to order a large candidate population by predicted fit. The structural parallel to wellbeing is exact: in both cases the analytic system produces an ordering of a population by predicted outcome, and its economic value derives from the ordering rather than from any therapeutic or developmental content of its own. Their study also records the conditions under which the ordering is reliable — data quality, representativeness of the training population, and continued human judgement in the decision — and those conditions transfer directly. A wellbeing risk model trained on the records of employees who chose to disclose, in an organisation where disclosure is unevenly safe, will order the population by willingness to disclose as much as by risk.

Two limits on the predictive claim are already visible in the assembled evidence and are developed quantitatively in Section V. The first is statistical: the precision with which a score identifies true cases depends on the prevalence of the condition in the screened population, and at the low prevalences characteristic of impending impairment in a general workforce, even a well-discriminating model returns a majority of false positives. The second is inferential: a model that predicts who will become impaired supplies no information about whether the intervention triggered by the prediction works. Prediction identifies; it does

not evaluate. An organisation that replaces retrospective cost accounting with predictive scoring, and does nothing else, has improved its timing while leaving its attribution problem exactly where it was.

2.7 The governance and trust constraint as an economic cost

Continuous wellbeing analytics operates on data that is health-adjacent when it is not health data outright, and the regulatory position in the United States is unusually unsettled. The incentive provisions of the Equal Employment Opportunity Commission's 2016 wellness regulations under the Americans with Disabilities Act and the Genetic Information Nondiscrimination Act were vacated following AARP v. EEOC and removed from the regulations with effect from 1 January 2019; replacement rules proposed in January 2021 were withdrawn before finalisation, leaving no operative federal incentive standard [18]. Employers therefore design incentive structures against the statutory requirement that participation be voluntary without an administrative safe harbour defining what voluntariness permits. State-level instruments — comprehensive consumer privacy statutes, health-specific legislation and biometric information statutes — have expanded into the resulting space unevenly, producing a compliance surface that varies by employee location within a single employer.

The economically important point is that the trust cost of surveillance is not an externality sitting outside the return calculation. It enters the calculation through the data. Employees who believe that disclosure of strain will be read as a signal of diminished reliability withhold the disclosure; utilisation of assistance programmes falls; the informative features degrade; and the model's discrimination declines precisely in the population where its value would be highest. A predictive system can therefore consume its own input, and the rate at which it does so is a function of the governance regime under which it operates.

Sastry and Arunachalam [19] address the general form of this problem in their treatment of probabilistic safe reinforcement learning, which models and manages risk in dynamic environments by constraining a learning system to operate within explicitly specified safety bounds while it optimises. Their formulation is directly transferable to workforce analytics and supplies the design principle this article adopts. A wellbeing prediction system is a learning system acting in an environment that reacts to it; leaving it to optimise on predictive accuracy alone permits it to reach states — eroded disclosure, adverse inference, discriminatory targeting — from which recovery is slow and the associated costs are not recorded in the objective function. The safe reinforcement learning approach of specifying the constraint set ex ante, and accepting a lower optimum within it, is the appropriate analogue for governance design in this domain, and Section VIII builds its policy recommendations on that principle.

2.8 Synthesis and research gap

The literature supports four propositions and leaves one gap. It supports the proposition that lost output is large, that its measured return to intervention declines as identification strengthens, that the returns which survive are those attached to targeted intervention on identified need, and that the recoverable margin lies substantially in the design of work rather than in individual health behaviour. The gap is that no study in the assembled set compares the return to intervention across measurement regimes while holding the intervention constant. The literature compares programmes; it does not compare the informational systems within which programmes are chosen and priced. This article addresses that gap by synthesis rather than by estimation, and specifies the design that would close it empirically.

III. RESEARCH METHODOLOGY

3.1 Design and rationale

The study adopts a systematic review with policy synthesis. No new estimation is undertaken and no causal parameter is claimed. The design is appropriate because the research questions concern the interpretation and comparability of an existing body of estimates rather than the generation of a further estimate, and because the central argument — that reported returns are a function of the measurement regime that produced them — can only be examined across studies. The review was conducted according to the reporting conventions of the PRISMA statement, adapted for an economic rather than clinical evidence base [29].

3.2 Search strategy and study selection

Searches were executed across Scopus, Web of Science, EconLit, PubMed, SSRN and the National Bureau of Economic Research working paper series, supplemented by targeted retrieval of grey literature from research institutes, federal agencies, employer coalitions and benefits research bodies. The search combined terms for the intervention (workplace wellness, wellbeing programme, employee assistance, health promotion), the outcome (productivity loss, presenteeism, absenteeism, lost output, return on investment) and the measurement approach (predictive analytics, people analytics, risk score, machine learning). Table 1 records the protocol.

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

Table 1: Systematic review protocol

Protocol element	Specification
Databases	Scopus; Web of Science; EconLit; PubMed; SSRN; NBER working papers
Grey literature	RAND; Integrated Benefits Institute; Business Group on Health; federal agency and employer-coalition reports
Period	January 2000 to June 2026
Population	Employed adults in the United States and comparable developed economies
Intervention	Employer-sponsored wellbeing, wellness, health-promotion or employee assistance programmes
Outcome	Any monetised or quantified measure of productivity loss, absence, presenteeism, turnover cost or programme return
Inclusion	Peer-reviewed or methodologically documented; comparator stated; extractable estimate or design detail
Exclusion	No economic outcome; no described method; vendor claim without stated sample or comparator; non-English
Coding	Identification tier; outcome type; reporting lag; treatment of selection; presenteeism measured or imputed

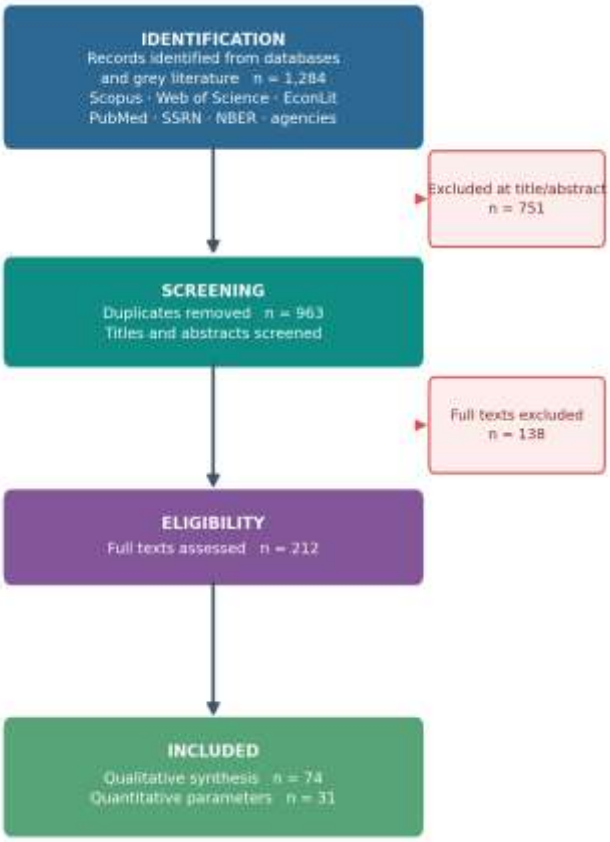


Figure 1: Study identification, screening and inclusion

Screening proceeded as shown in Figure 1. Of 1,284 records identified, 963 remained after removal of duplicates, 212 proceeded to full-text assessment, and 74 met the inclusion criteria for qualitative synthesis, of which 31 supplied quantitative parameters in a form permitting extraction. Vendor and consultancy material was retained only where sample, comparator and estimation approach were stated; material reporting a return without a described method was excluded, and 33 records were excluded on that ground alone.

3.3 Quality appraisal and the evidence hierarchy

Included studies were classified by identification strategy into four tiers: randomised trials with pre-specified outcomes; quasi-experimental designs with a credible comparison group; observational designs comparing participants with non-participants; and modelled or actuarial estimates without an empirical comparator. This classification is the analytical spine of the review, because the article's central finding concerns the relationship between tier and reported return. Studies were

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

additionally coded for the outcome measured, the reporting lag between exposure and measurement, the treatment of selection, and whether presenteeism was measured or imputed.

3.4 Synthesis approach

Findings were synthesised into a two-regime framework distinguishing retrospective cost measurement from predictive wellbeing analytics along seven dimensions: detection lead time, causal attribution, targeting precision, coverage of presenteeism, privacy and trust cost, governance maturity and operating cost. Studies were mapped to the regime whose informational properties they instantiate, and the return to intervention was then compared across regimes.

3.5 Status of the decision-analytic illustrations

Sections IV and V include arithmetic illustrations — the decay of recoverable loss with detection lag, and the relationship between prevalence, predictive value and cost per case identified. These are deterministic computations on parameters drawn from the reviewed literature, presented to make the structure of the argument explicit. They are not estimates, carry no confidence intervals, and should not be read as findings about any particular employer. Their sole function is to display the shape of the relationships that the reviewed evidence implies.

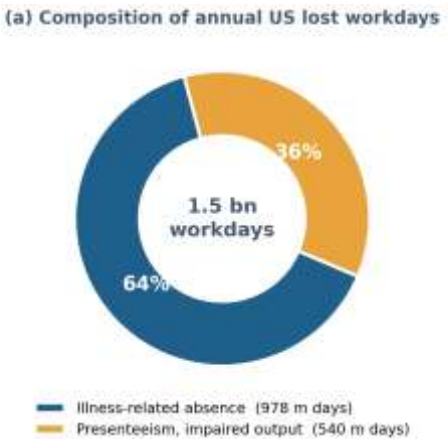
IV. FINDINGS: THE RETROSPECTIVE MEASUREMENT REGIME

4.1 What the regime observes and what it does not

Table 2 maps the components of the cost of lost output against the instrument by which each is captured, the lag with which it is reported and the principal source of measurement error. The pattern that emerges is systematic rather than incidental: measurement precision is inversely related to component size. Direct medical claims, which are the smallest component of the economic loss, are measured with administrative accuracy. Presenteeism, which the Integrated Benefits Institute decomposition places at 540 million days annually [1], is measured by annual self-report with recall and social-desirability error and a demonstrated tendency of respondents to under-report impairment where reporting could be read as a performance signal.

Table 2: Components of the cost of lost output and their measurement properties

Component	Instrument	Reporting lag	Principal error
Medical and pharmacy claims	Administrative claims	3-6 months to run-out; annual analysis	Low; administratively accurate
Sickness absence	Payroll and leave register	1-3 months	Low for recorded absence; misses informal absence
Disability and workers' compensation	Insurer records	6-18 months	Low on payment; weak on causation
Presenteeism	Annual self-report scale (HPQ, WLQ, SPS-6)	12 months or more	High; recall, social desirability, performance-signal suppression
Wellbeing-related turnover	Exit interview; attrition analysis	Post hoc; often unattributed	High; attribution to wellbeing rarely established
Impaired team output	Not routinely measured	Not observed	Component largely absent from the accounting



EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

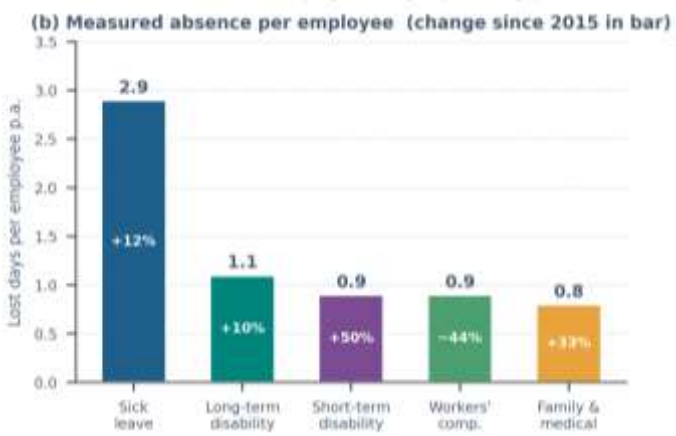


Figure 2: Composition of measured lost output in United States employment (Integrated Benefits Institute benchmarking data)

The consequence is a systematic misallocation of programme design effort. Because the claims margin is legible and the presenteeism margin is not, wellbeing programmes are designed, justified and evaluated against the smaller quantity. This is a textbook instance of optimisation against the measured proxy rather than the objective, and it predicts exactly the pattern observed in the randomised trials, where programmes produced modest movement in self-reported behaviour and none in the economic outcomes the employer cares about.

4.2 Attribution failure

Figure 3 arranges the principal return estimates in the assembled evidence by the identification strategy that produced them. The ordering is monotonic. Observational meta-analysis yields 3.27:1 on medical cost and 2.73:1 on absenteeism [8]; quasi-experimental employer data yield 1.50:1 overall [9]; two randomised trials yield no statistically detectable effect on spending, absence, tenure or performance [10, 12]. Table 3 records the corresponding study characteristics.

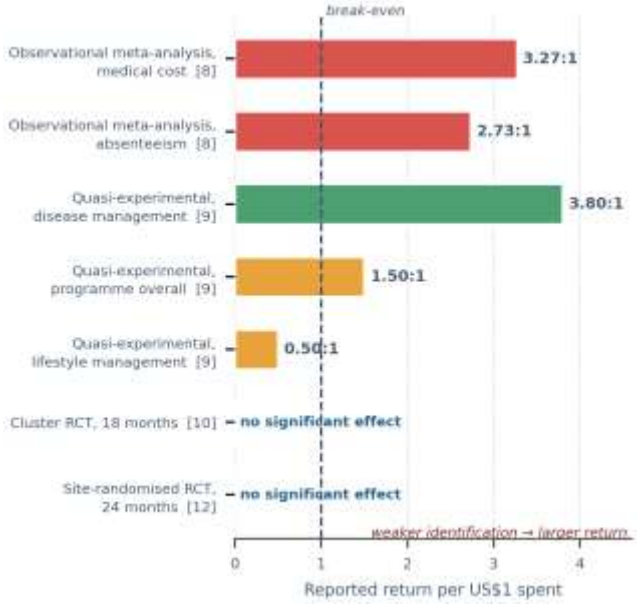


Figure 3: Reported return to wellbeing expenditure by strength of identification

Table 3: Principal included studies contributing return estimates

Study	Design	Scale	Reported effect on economic outcomes
Baicker, Cutler and Song [8]	Meta-analysis, largely observational	22 studies	US$3.27 medical and US$2.73 absenteeism saved per US$1
Mattke and others, RAND [9]	Quasi-experimental employer data	Large multi-employer panel	1.50:1 overall; 3.80:1 disease management; 0.50:1 lifestyle management; about US$30 PMPM
Song and Baicker [10]	Cluster randomised trial, 18 months	32,974 employees; 160 worksites	2 of 80 outcomes significant, both self-reported behaviour; no effect on spending, absence, tenure or performance

Song and others [11]	Same trial, 3-year follow-up	As above	Behavioural differences persist; no effect on clinical measures, spending or employment outcomes
Jones, Molitor and Reif [12]	Site-randomised trial, 24 months	About 4,800 employees	Screening participation up; no effect on spending, behaviour, productivity or health; CIs exclude about 84% of prior estimates
Chisholm and others [14]	Global return-on-investment model	36 countries, 2016-2030	Benefit-cost 2.3-3.0:1 economic; 3.3-5.7:1 including health returns
Deloitte [31]	Employer survey with modelled cost	UK employers, 3,599 respondents	Total cost GBP 53-56 bn; average GBP 5 returned per GBP 1; proactive designs outperform reactive

The economic reading of this gradient is that the reported return under the retrospective regime is not a measurement of programme effect but a composite of programme effect and selection, in which the second term dominates. The Illinois baseline finding — that future participants were already cheaper and healthier before the programme existed [12] — identifies the composite directly. For capital allocation this is more damaging than a null result would be, because a null result is informative while a biased estimate of unknown magnitude is not. An employer presented with a 3:1 return computed on participant-versus-non-participant comparison cannot determine from the number whether the true return is 3:1, 1:1 or below break-even.

4.3 Latency and the decay of recoverable value

The second pathology is temporal. Retrospective measurement observes a loss after it has been incurred, at which point the output is unrecoverable and only its recurrence is addressable. The share of a loss episode that remains avertible therefore declines with the lag between onset and detection. Figure 4 displays this relationship for the reporting cadences documented in the reviewed literature, using an exponential decay calibrated so that half the avertible value is lost at approximately six months.

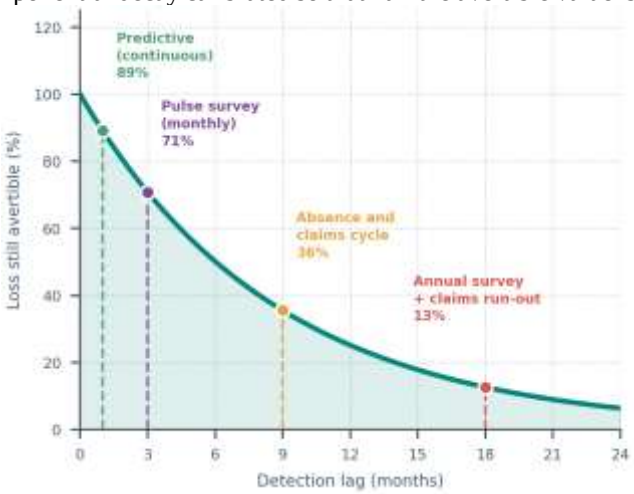


Figure 4: Recoverable share of an output-loss episode by detection lag, with reporting cadences of common measurement instruments

Under an annual engagement survey combined with claims run-out, at a lag of roughly eighteen months, approximately thirteen per cent of the episode value remains avertible. Under an absence and claims reporting cycle at nine months the figure is about thirty-six per cent. Under continuous operational monitoring at a lag of one month it is approximately eighty-nine per cent. The specific parameterisation is illustrative, but the qualitative conclusion is robust to any monotone decay function: the retrospective regime forfeits most of the avertible value before it produces a signal, and it does so structurally rather than through poor execution.

4.4 Denominator error

The third pathology follows from the first two. Because presenteeism is imputed rather than observed, and because it is the larger component of impaired output, the denominator of any return calculation carries an error term of unknown sign and material magnitude. Programmes that reduce presenteeism without reducing claims will appear ineffective; programmes that shift cost between benefit categories without reducing total impairment may appear effective. Neither error is detectable within the regime that generates it.

V. FINDINGS: THE PREDICTIVE MEASUREMENT REGIME

5.1 What prediction changes

Predictive wellbeing analytics changes two things and, on the evidence assembled here, only two. It reduces detection lag from months to days or weeks, which moves the employer up the recoverable-value curve of Figure 4. And it orders the population by predicted risk, which permits intervention to be concentrated on the segment where expected loss is highest rather than distributed uniformly. The second of these is the economically decisive change, because it manufactures the condition —

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

targeting on identified need — that distinguishes the interventions which survived randomised evaluation from those which did not.

This is a more precise and more modest claim than the one usually advanced for these systems. Predictive analytics does not make an ineffective intervention effective. It raises the yield of an effective intervention by delivering it earlier and to a better-selected population. Where no effective intervention exists for the identified risk, the analytic layer generates cost and no return, and where the effective intervention is an organisational change rather than an individual service, the score must be actionable at organisational level for any return to be realised at all.

5.2 The base-rate constraint on targeting value

The value of targeting is bounded by the precision with which the target is identified, and that precision depends on prevalence as well as on model quality. For a classifier with sensitivity Se and specificity Sp applied to a population with prevalence p, the positive predictive value is Se·p divided by the quantity Se·p plus (1−Sp)·(1−p). Figure 5 traces this function, and the associated outreach cost per true case identified, across the prevalence range relevant to impending workplace impairment.

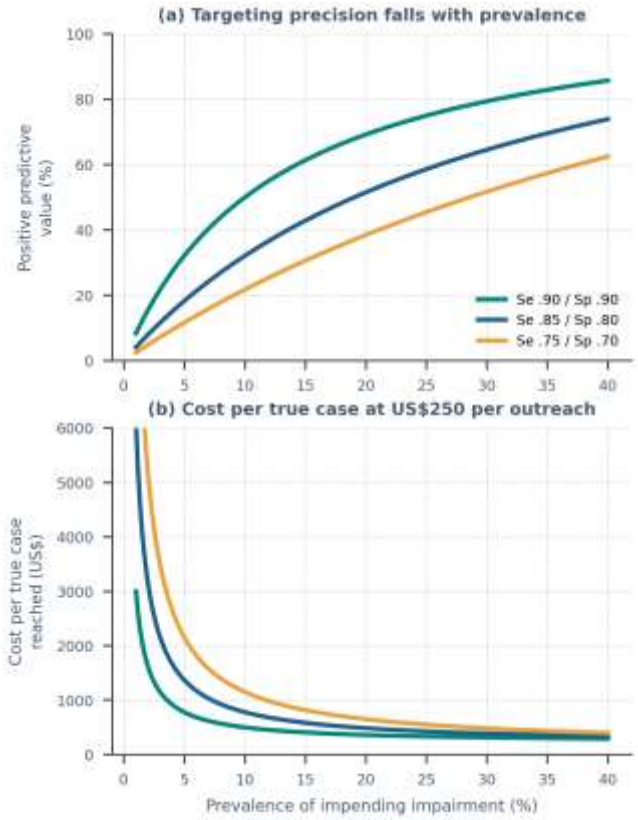


Figure 5: Positive predictive value and outreach cost per true case identified, by prevalence and classifier performance

Table 4: Positive predictive value and outreach cost per true case identified

Prevalence	PPV at Se .90 / Sp .90	PPV at Se .85 / Sp .80	Cost per true case (Se .85 / Sp .80)
2%	0.155	0.080	US$3,133
3%	0.218	0.116	US$2,152
5%	0.321	0.183	US$1,368
10%	0.500	0.321	US$779
15%	0.614	0.429	US$583
20%	0.692	0.515	US$485
30%	0.794	0.646	US$387

Table 4 records the arithmetic. A classifier with sensitivity of 0.85 and specificity of 0.80 — respectable performance for a model built on operational rather than clinical features — achieves a positive predictive value of about 0.32 at a prevalence of ten per cent, so that roughly two of every three employees contacted are not in fact at risk. At a prevalence of three per cent the same model achieves a positive predictive value near 0.12, and at an outreach cost of US$250 per flagged employee the cost

per true case reached rises above US$2,000. Improving the model helps but does not escape the constraint: raising both parameters to 0.90 lifts predictive value at three per cent prevalence only to about 0.22.

Three practical implications follow. First, prevalence is a design variable, not merely a fact of nature: applying a model to a pre-identified high-risk stratum — a department with elevated overtime, a cohort in a restructuring unit — raises the prevalence in the screened population and therefore the precision of the score, which is the analytic counterpart of the disease-management result [9]. Second, the cost of a false positive is not zero even when the intervention offered is benign, because unnecessary contact on a wellbeing basis carries an inferential cost for the recipient that feeds directly into the trust dynamic of Section 2.7. Third, at low prevalence, intervention designs whose cost does not scale with the number of individuals contacted — workload rebalancing, schedule redesign, managerial capability building — dominate designs whose cost does.

5.3 Prediction is not evaluation

The most consequential limitation of the predictive regime is that it inherits the attribution problem of the retrospective regime intact. A risk score identifies employees or units likely to experience impairment. If those identified receive an intervention and their subsequent outcomes are compared with the outcomes of those not identified, the comparison is confounded by the very risk differential the model was built to detect. If they are compared with their own predicted trajectory, the comparison is confounded by regression to the mean, which is severe precisely where selection is on an extreme predicted value. Neither comparison identifies a causal effect. Predictive systems can therefore reproduce the inflated returns of the observational literature with greater technical sophistication and greater persuasiveness.

This limitation is not intrinsic. It is a consequence of deploying prediction without an evaluation design, and it is removed at low cost by randomising the intervention among those the model flags. Where all flagged employees cannot be served immediately, which is the ordinary condition, randomised phasing of outreach across flagged individuals or stepped-wedge rollout across flagged units produces an unbiased estimate of the intervention's effect at no additional programme cost and with no denial of service to anyone who would otherwise have received it. The predictive regime is thus the first measurement architecture in this field that makes credible evaluation cheap, and the failure of current implementations to exploit that is the clearest actionable finding of this review.

VI. COMPARATIVE RETURN TO INTERVENTION UNDER THE TWO REGIMES

Table 5 sets the two regimes against each other on the seven dimensions of the synthesis framework, and adds the hybrid architecture that the foregoing analysis implies. Figure 6 displays the same comparison graphically.

Table 5: The two measurement regimes and the proposed hybrid architecture

Dimension	Retrospective cost measurement	Predictive wellbeing analytics	Predictive targeting with prospective randomised evaluation
Detection lead time	Negative: loss observed after it is incurred, 9-24 months	Days to weeks; continuous	Days to weeks; continuous
Causal attribution	Weak: selection dominates participant comparisons	Weak: regression to the mean and risk confounding persist	Strong: randomised phasing among flagged units identifies the effect
Targeting precision	Low: uniform offer to the covered population	Moderate to high, bounded by base rate	High: prediction used for allocation, not for evaluation
Coverage of presenteeism	Low: annual self-report only	Higher: continuous operational proxies	Higher, with periodic validated instrument for calibration
Privacy and trust cost	Low: data already collected for benefit administration	High: continuous observation; risk of disclosure chilling	Moderate: organisational features preferred over individual health data
Governance maturity	High: settled treatment and established practice	Low: standards unsettled; inference from non-health data unresolved	Moderate: constraint set specified in advance
Operating cost	Low: incremental to existing benefit administration	Moderate to high: platform, integration and analytic capability	Moderate: marginal cost of randomised phasing is close to zero

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

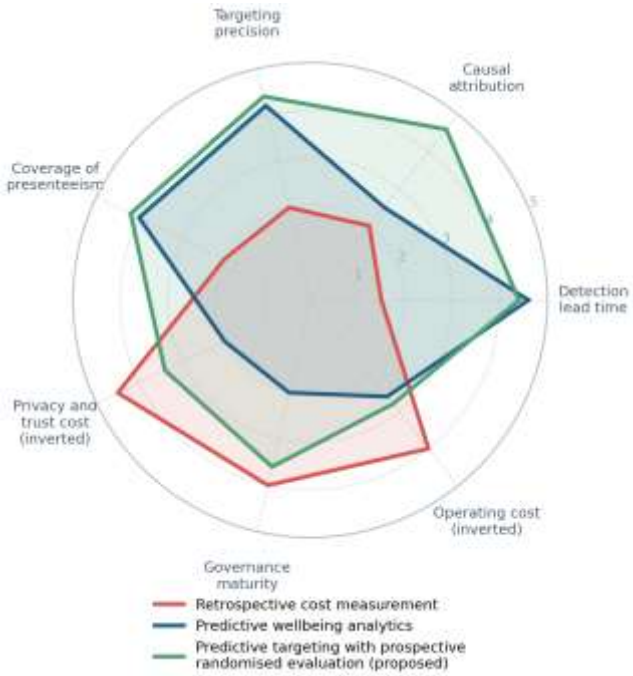


Figure 6: Comparative properties of the two measurement regimes and the proposed hybrid architecture

The comparison yields a result that neither advocacy position anticipates. The retrospective regime is superior on exactly two dimensions — governance maturity and operating cost — and both advantages are transitional rather than structural. Governance maturity is higher because the regime has existed longer and its legal treatment is settled; operating cost is lower because the data are already collected for other purposes. The predictive regime is superior on detection lead time, targeting precision and coverage of presenteeism, and inferior on privacy and trust cost. On causal attribution, which is the dimension that determines whether a reported return can be relied upon for capital allocation, both regimes score poorly, and the predictive regime scores only marginally better than the retrospective one.

The economically important conclusion follows directly. The choice between retrospective and predictive measurement is not the choice that determines whether the return to wellbeing intervention can be established. Both regimes, deployed as they ordinarily are, leave attribution unresolved. What determines it is whether an evaluation design is embedded in the measurement architecture, and the predictive regime's distinctive advantage is that it makes such embedding nearly costless. The hybrid architecture in Figure 6 is therefore not a compromise between the two positions but a specification of the conditions under which the forward-looking regime delivers the return its advocates claim for it.

Table 6 renders that specification as a comparison of what each regime can and cannot support in the way of managerial decision, which is the form in which the question presents itself to an employer choosing between them.

Table 6: Managerial decisions each regime can and cannot support

Decision	Retrospective	Predictive	Hybrid
Reconcile benefit spend to the financial statements	Yes	No	Yes
Identify which units are accumulating output loss	With long lag	Yes	Yes
Decide whom to reach, and when	No	Yes	Yes
Establish whether the intervention caused the change	No	No	Yes
Price the return for capital allocation	No	No	Yes
Support a claim of programme effectiveness to a third party	Not defensibly	Not defensibly	Yes

VII. THE INDIAN COMPARISON

The analysis to this point has been set in the United States and comparable developed economies, where the employer is the residual claimant on health-related productivity loss because it also finances health care. India presents a different institutional configuration, and the comparison is instructive in both directions: it clarifies which features of the argument are general and

EPRA International Journal of Socio-Economic and Environmental Outlook (SEEO) ISSN: 2348-4101
Volume: 13 | Issue:9 |Setember 2026 | SJIF Impact Factor (2026): 8.701 | Journal DOI: 10.36713/epra0314 |Peer-Reviewed Journal

which are artefacts of American benefit financing, and it identifies a set of conditions under which the predictive regime may be adopted in India without first passing through the retrospective one.

7.1 Scale and composition

The Indian evidence base on lost output is thinner than the American, and is concentrated on mental health. Deloitte's survey of the Indian workforce placed the annual cost of poor employee mental health to Indian employers at approximately US$14 billion, arising jointly from absenteeism, presenteeism and attrition, and found that eighty per cent of respondents reported mental health symptoms during the preceding year, with forty-seven per cent identifying workplace stress as the principal contributor [20]. The behavioural distribution in that survey is directly relevant to the measurement argument: thirty-three per cent of affected respondents continued working while unwell, twenty-nine per cent took time off, twenty per cent resigned, and thirty-nine per cent avoided taking any step at all because of stigma. Read as a measurement statement rather than a wellbeing statement, this says that the modal Indian response to impairment is the one that leaves no administrative trace. The presenteeism share of Indian lost output is therefore likely to exceed the American share, while being measured less often.

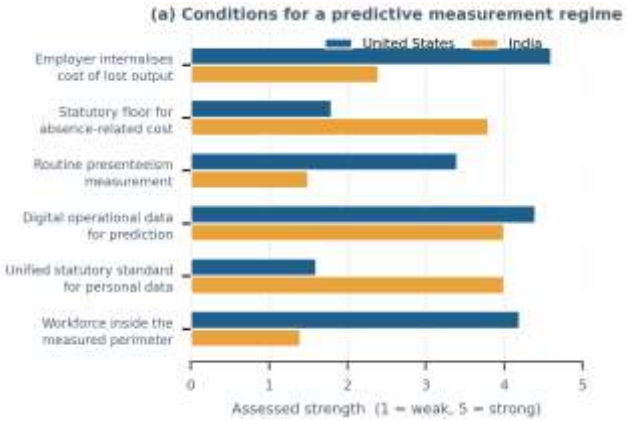
At the macroeconomic level the World Economic Forum and Harvard School of Public Health projected cumulative Indian output losses of US$4.58 trillion from non-communicable diseases and mental health conditions over 2012 to 2030, with mental health conditions among the largest single contributors [21]. The order of magnitude confirms that the economic category is not smaller in India than in the developed economies; what differs is who bears it and who can observe it.

7.2 Institutional divergence: who bears the cost

Three structural features separate the Indian case. First, the statutory floor. The Employees' State Insurance scheme covers employees in establishments of ten or more workers earning up to ₹21,000 per month, and pays sickness benefit at seventy per cent of average daily wages for up to ninety-one days across two consecutive benefit periods, subject to a contribution condition of seventy-eight days [22]. For the covered population, a substantial part of the direct cost of sickness absence is socialised rather than borne by the employer, which weakens the private incentive to invest in prevention relative to the American case even as it strengthens the social return to doing so.

Second, the wage ceiling partitions the organised sector. Above ₹21,000 per month, employers self-insure through group medical cover, and the American configuration reasserts itself in attenuated form: the employer bears claims cost and therefore has a legible, if partial, cost signal. Below it, the employer bears the output loss but not the medical cost, so the only quantity it can observe is absence — the component this article has shown to be the smaller and less economically significant one.

Third, and decisively, the informal sector lies outside both perimeters. The great majority of Indian workers are informally employed [30], with neither statutory sickness benefit nor employer group cover, and consequently with no administrative record of health-related output loss of any kind. Figure 7(b) renders the resulting structure. For this population the cost of lost output is real, is borne by the worker and the household as forgone earnings, and is invisible to every instrument discussed in this article. Any Indian estimate of the national cost of impaired working time built from employer records is therefore a lower bound on a small subset, and should not be compared directly with American aggregates such as the Integrated Benefits Institute series [1].



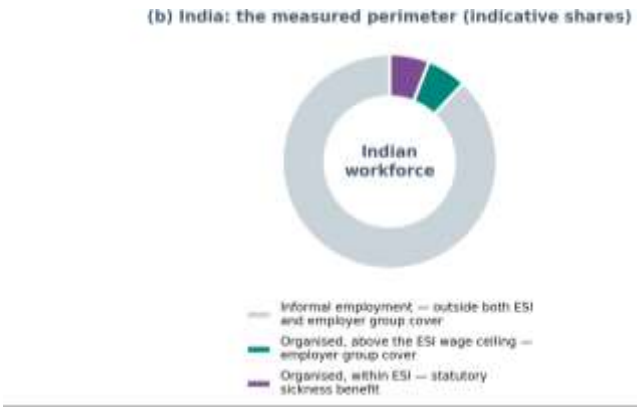


Figure 7: Comparative conditions for a predictive measurement regime, and the structure of the measured perimeter in India

7.3 Leapfrog or replicate?

Figure 7(a) assesses the two economies against the conditions this article has identified as determining whether a predictive regime can be operated productively. The pattern is not one of uniform Indian disadvantage. India scores lower on employer internalisation of the cost of lost output, on established presenteeism measurement practice and, very much lower, on the share of the workforce inside any measured perimeter. It scores higher on the presence of a statutory floor for absence-related cost and on the existence of a unified statutory standard for personal data. On the availability of digital operational data — the raw material of prediction — the two are close, because India's organised knowledge-services sector runs on the same workflow, scheduling and collaboration platforms as its American counterpart and generates the same exhaust.

The implication is that the organised Indian sector is in a position to adopt the predictive regime without first building the retrospective apparatus that the United States spent three decades constructing and that this article has shown to be a poor guide to intervention. The claims-data infrastructure whose legibility drew American programme design toward the smaller cost margin is largely absent in India below the wage ceiling. That absence is conventionally treated as a deficiency; on the argument developed here it is closer to an opportunity, because it removes the proxy against which American programmes have been mis-optimised. An Indian employer building a wellbeing measurement system today faces a genuine choice of architecture rather than an inherited one.

This conclusion is strengthened by the observation that the evidence locating the recoverable margin in job design rather than individual health behaviour is itself substantially Indian in origin. The studies establishing that delegated authority requires matched resource provision to raise rather than depress effectiveness [15, 16], that burnout is produced by leadership practice and workload allocation rather than by individual health behaviour [13], and that task achievement varies with departmental configuration [4], were all conducted in Indian workplaces. The organisational features they identify — workload, span of control, decision latitude, resource adequacy — are observable in Indian human resource information systems without any health data being processed at all, which places the most evidentially supported intervention target inside the least regulated data category.

7.4 Governance: a unified standard against a regulatory vacuum

The governance comparison inverts the usual expectation. The United States, as Section 2.7 recorded, has operated without a federal wellness incentive standard under the Americans with Disabilities Act and the Genetic Information Nondiscrimination Act since the vacatur of the 2016 provisions, with a state-level patchwork filling the space unevenly [18]. India, by contrast, now has a single consent-based statute of general application: the Digital Personal Data Protection Act, 2023, whose implementing Rules were notified on 14 November 2025 with phased compliance, requires informed, unambiguous and freely given consent, a specific-purpose notice, purpose limitation, retention limits and a withdrawal mechanism, with additional obligations on significant data fiduciaries [23].

Two consequences follow for wellbeing analytics in India. The first is favourable: an employer deploying a predictive system has a single, ascertainable standard to design against rather than a compliance surface that varies by employee location. The second is a genuine difficulty. Indian data protection law does not carve out a separate heightened category for health or biometric data in the manner of the earlier sensitive personal data rules or of comparable European provisions; it regulates all personal data through the consent architecture. Where consent is the sole gate, the voluntariness of consent in an employment relationship — a relationship characterised by asymmetric bargaining power — becomes the load-bearing element of the entire protective structure. This is precisely the question that the vacated American incentive rules attempted, unsuccessfully, to settle.

Alongside this, the four labour codes brought into force on 21 November 2025 consolidated twenty-nine existing statutes and extended social security architecture toward gig and platform workers, with aggregator contributions of one to two per cent of annual turnover directed to worker welfare funds [24]. The Occupational Safety, Health and Working Conditions Code within that set moves occupational health from a fragmented factory-era framework toward a general employer duty. For the present

argument the significance is that the perimeter of measurable employment in India is being enlarged by statute at the same moment that the analytic capacity to measure within it is arriving, which is a conjunction the developed economies did not enjoy.

7.5 What transfers and what does not

Table 7 sets out the comparison in full. Three elements of the argument transfer without modification. The attribution result transfers: a participant-versus-non-participant comparison is contaminated by selection in Bengaluru exactly as it is in Illinois, and Indian employers reporting such returns should be read with the same scepticism. The latency result transfers: the decay of recoverable value with detection lag is a property of impairment, not of institutions. The base-rate constraint on targeting transfers, and binds somewhat harder, because the prevalence of formally identified impairment is lower where stigma suppresses disclosure [20], which depresses positive predictive value at any given model quality.

Two elements do not transfer. The private return to wellbeing investment is lower in India for the covered organised population, because sickness cost is partly socialised through the Employees' State Insurance scheme, so the case for employer investment must rest more heavily on output and retention than on claims avoidance — which, given that this article has argued the claims margin to be the wrong target in any case, is not the disadvantage it first appears. And the policy conclusion for regulators differs in kind: the American requirement is to restore a standard that has lapsed, whereas the Indian requirement is to specify how a general consent statute applies to a relationship in which consent is structurally constrained. The second is the harder problem, and it is being confronted at a more favourable moment, before the practices it must govern have hardened.

Table 7: Comparative structure of the problem in the United States and India

Dimension	United States and comparable developed economies	India
Who finances health care for the workforce	Predominantly the employer, which makes it the residual claimant on health-related output loss	Statutory Employees' State Insurance below the wage ceiling; employer group cover above it; nothing for informal employment
Statutory absence cost floor	No general federal paid sick leave entitlement; provision varies by state and employer	Sickness benefit at 70% of average daily wages for up to 91 days in two consecutive benefit periods [22]
Headline cost estimate	About US$575 billion and 1.5 billion lost workdays annually, all causes of poor health [1]	About US$14 billion annually to employers from poor mental health alone [20]; US$4.58 trillion cumulative NCD and mental health loss 2012-2030 [21]
Share of workforce inside the measured perimeter	High: most employment is formal and administratively recorded	Low: the large majority of workers are informally employed and generate no record of health-related output loss [30]
Dominant unmeasured component	Presenteeism, imputed annually by self-report	Presenteeism, rarely measured at all; 33% of affected employees report working through illness and 39% take no step because of stigma [20]
Data protection standard	No operative federal wellness incentive standard since 2019; state patchwork [18]	Single consent-based statute of general application: DPDP Act 2023 with Rules notified 14 November 2025 [23]
Heightened category for health data	Yes, through ADA and GINA disability-inquiry and genetic-information constraints	No separate category; all personal data governed through the consent architecture, so voluntariness of employee consent is load-bearing
Direction of institutional change	Standard lapsed and not restored; regulatory attention shifting to algorithmic management	Perimeter widening: four labour codes in force from 21 November 2025 extend social security toward gig and platform work [24]
Implication for measurement architecture	Retrospective apparatus already built; the task is to add prospective evaluation to it	Retrospective apparatus largely absent below the ceiling; the organised sector can adopt predictive targeting with embedded evaluation directly

VIII. POLICY SYNTHESIS

8.1 For employers

Employers should cease reporting participant-versus-non-participant return estimates, internally or externally, since such estimates are known to be dominated by selection [12] and their circulation degrades the quality of decisions elsewhere in the market. Where a predictive capability is adopted, randomised phasing of outreach among flagged individuals or stepped-wedge rollout across flagged units should be built into the deployment from the outset rather than added later, because retrofitting an

evaluation design to a live programme is substantially harder than specifying one before launch. Risk features should be drawn preferentially from organisational and operational data — workload, span of control, schedule volatility, resource adequacy, decision latitude — rather than from biometric or claims sources, on three grounds: this is where the evidence locates the recoverable margin [13, 15, 16]; these variables are directly manipulable by management whereas individual health status is not; and they avoid the regulatory and trust exposure that health data carries. Intervention should be concentrated on identified units at elevated risk rather than distributed uniformly, which is the operational form of the targeting result [9, 14].

8.2 For insurers and benefit intermediaries

Carriers and benefit consultancies occupy the position in this market where measurement standards could be set at low cost, since they supply the analytic infrastructure to employers too small to build it. Three measures follow from the review. Reported returns should be accompanied by mandatory disclosure of the identification strategy that produced them, using a standard tier classification of the kind adopted in Section 3.3, so that a purchaser can distinguish a randomised estimate from a participant comparison. Presenteeism should be measured with validated instruments [5] at a cadence closer to the rate at which impairment develops than to the enrolment calendar. And vendor contracts should specify evaluation design as a deliverable rather than treating it as a discretionary extra, since the marginal cost of randomised phasing within an already-funded programme is close to zero.

8.3 For regulators

The absence of an operative federal incentive standard under the Americans with Disabilities Act and the Genetic Information Nondiscrimination Act [18] is now a sustained rather than a transitional condition, and it is economically costly in a way that is not widely recognised. Uncertainty about permissible incentive design suppresses investment by cautious employers while leaving aggressive designs unconstrained, which is the opposite of the intended distribution. Restoring a clear incentive standard would reduce compliance cost and narrow the variance of programme design. Beyond incentives, the expansion of predictive analytics into health-adjacent operational data indicates a second requirement: a clear rule on when inference from non-health data to health status constitutes a disability-related inquiry, since the current position permits an employer to derive by model what it may not ask directly.

The design principle for such rules is the one Sastry and Arunachalam [19] articulate for learning systems generally: specify the constraint set in advance and require the optimising system to operate inside it, accepting a lower attainable optimum in exchange for the exclusion of states from which recovery is slow. Applied here, this argues for ex ante limits on the feature classes admissible to wellbeing models and on the decisions to which their outputs may be applied, rather than ex post adjudication of harms after a model has been deployed at scale.

8.4 For the research community

The gap identified in Section 2.8 is addressable with existing data. The design required is a trial in which the intervention is held constant while the measurement and targeting regime is randomised at unit level, so that the return to intervention under retrospective and predictive targeting can be compared directly. Employers adopting predictive systems are in a position to run such trials at negligible marginal cost, and the principal obstacle is not methodological but institutional: the results are commercially sensitive and there is at present no mechanism through which they enter the public evidence base. A pre-registration facility for employer wellbeing evaluations, on the model of clinical trial registration, would address the resulting publication asymmetry, which currently biases the visible literature toward favourable findings from weak designs.

IX. LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Four limitations qualify these findings. The systematic review is confined to the United States and comparable developed economies, where employer-sponsored health financing makes the employer the residual claimant on health-related productivity loss; the Indian comparison in Section VII is a structured institutional analysis rather than a second systematic review, and the Indian empirical base on lost output is thin enough that its figures should be treated as orders of magnitude rather than as estimates. The evidence base is dominated by large employers with formal benefit administration, and the small-employer sector, which accounts for a substantial share of employment, is largely unobserved. The randomised evidence that anchors the attribution argument derives from two trials of individually-oriented programmes, so the null result is strictly a null for that class of intervention and not for organisational redesign, on which randomised evidence remains sparse. And the decision-analytic illustrations in Sections IV and V are deterministic computations on review-derived parameters, presented to display structure and carrying no claim to estimate any employer's actual return.

Four lines of further work follow. The regime-randomised trial specified in Section 8.4 is the priority. Second, randomised evaluation of organisational interventions — workload rebalancing, span-of-control adjustment, schedule stabilisation, resource provision matched to delegated authority [15, 16] — would test directly the proposition that the recoverable margin lies in job design. Third, the trust dynamic of Section 2.7 is empirically open: the elasticity of disclosure with respect to perceived surveillance intensity is the parameter that determines whether predictive systems degrade their own inputs, and it has not been estimated. Fourth, the Indian propositions of Section VII are testable: whether organised-sector employers that adopt predictive targeting without a prior claims apparatus achieve better-identified returns than their American counterparts is an empirical question, and answering it would establish the boundary conditions of the framework.

X. CONCLUSION

The return to employee wellbeing investment has proved unstable not because the underlying economic loss is small — it is very large — but because the measurement regime through which the return is calculated cannot separate effect from selection, observes losses only once they are unrecoverable, and measures its largest cost component least precisely. The reported return falls monotonically as identification strengthens, from above 3:1 in observational work to no detectable effect in randomised trials whose confidence intervals exclude most prior estimates. The returns that survive rigorous evaluation share one feature, which is that intervention was targeted on identified need.

Predictive wellbeing analytics is best understood, economically, as a targeting technology rather than a therapeutic one. Its value derives from reducing detection lag and from ordering the population by predicted risk, and that value is bounded by the prevalence of the risk in the screened population, by the efficacy of the interventions the score triggers, and by a privacy and trust cost that enters the calculation through its effect on disclosure and hence on the quality of the model's own inputs. Deployed without an evaluation design, the predictive regime reproduces the attribution failure of the regime it replaces, in a more sophisticated and more persuasive form.

The measurement regimes are therefore not substitutes but complements arranged in the wrong order. Prediction should perform the targeting, for which retrospective measurement is structurally unsuited; prospective randomisation embedded in the predictive workflow should perform the evaluation, for which prediction is equally unsuited; and retrospective accounting should be retained for the financial reconciliation at which it is genuinely reliable. Under that architecture the return to wellbeing intervention becomes an estimable quantity rather than a contested one, and the cost of lost output becomes a margin an employer can manage rather than a figure it can only report.

REFERENCES

[1] *Integrated Benefits Institute. (2021). Poor health costs US employers $575 billion and 1.5 billion days of lost productivity. IBI Research Insights, Oakland, CA.*
[2] *Gallup. (2026). State of the Global Workplace: The Voice of the World's Employees. Gallup Inc., Washington, DC.*
[3] *Sastry, N. S. K., & Mallya, M. M. (2025). Strategic HRM for economic resilience: Skill diversification, labor market dynamics and technology disruption. EPRA International Journal of Research and Development, 10(11), 408-414.*
[4] *Sastry, N. S. K. (2024). An impact of departmental structure on task achievement. EPRA International Journal of Environmental, Economics, Commerce and Educational Management, 11(9).*
[5] *Kessler, R. C., Barber, C., Beck, A., Berglund, P., Cleary, P. D., McKenas, D., Pronk, N., Simon, G., Stang, P., Ustun, T. B., & Wang, P. (2003). The World Health Organization Health and Work Performance Questionnaire (HPQ). Journal of Occupational and Environmental Medicine, 45(2), 156-174.*
[6] *Business Group on Health. (2025). 2025 Employer Well-being Strategy Survey. Business Group on Health, Washington, DC.*
[7] *Sastry, N. S. K., & Selvarasu, A. (2014). A study of impact on performance appraisal on employee's engagement. International Journal of Managerial Studies and Research, 2(11), 10-22.*
[8] *Baicker, K., Cutler, D., & Song, Z. (2010). Workplace wellness programs can generate savings. Health Affairs, 29(2), 304-311.*
[9] *Mattke, S., Liu, H., Caloyeras, J. P., Huang, C. Y., Van Busum, K. R., Khodyakov, D., & Shier, V. (2013). Workplace Wellness Programs Study: Final Report. RAND Corporation, RR-254-DOL, Santa Monica, CA.*
[10] *Song, Z., & Baicker, K. (2019). Effect of a workplace wellness program on employee health and economic outcomes: A randomized clinical trial. JAMA, 321(15), 1491-1501.*
[11] *Song, Z., Baicker, K., & colleagues. (2021). Health and economic outcomes up to three years after a workplace wellness program: A randomized controlled trial. Health Affairs, 40(6), 951-960.*
[12] *Jones, D., Molitor, D., & Reif, J. (2019). What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study. Quarterly Journal of Economics, 134(4), 1747-1791.*
[13] *Sastry, N. S. K. (2024). Revitalizing leadership to combat workforce burnout. EPRA International Journal of Environmental, Economics, Commerce and Educational Management, 11(9).*
[14] *Chisholm, D., Sweeny, K., Sheehan, P., Rasmussen, B., Smit, F., Cuijpers, P., & Saxena, S. (2016). Scaling-up treatment of depression and anxiety: A global return on investment analysis. The Lancet Psychiatry, 3(5), 415-424.*
[15] *Sastry, N. S. K., & Ravish, P. Y. (2025). Empowering employees in the Indian workplace: The transformative power of delegation and resources. EPRA International Journal of Research and Development, 10(7).*
[16] *Sastry, N. S. K., & Ravish, P. Y. (2025). The impact of delegated authority and resource provision on decision-making: Fostering employee initiative. EPRA International Journal of Multidisciplinary Research, 11(7).*
[17] *Sastry, N. S. K., & Mallya, M. M. (2025). Revolutionizing HR hiring with AI analytics: Enhancing efficiency and attracting top talent. EPRA International Journal of Research and Development, 10(9), 125-131.*
[18] *U.S. Equal Employment Opportunity Commission. Regulations under the Americans with Disabilities Act and the Genetic Information Nondiscrimination Act on employer wellness programs; incentive provisions vacated following AARP v. EEOC, 267 F. Supp. 3d 14 (D.D.C. 2017), removed with effect from 1 January 2019; proposed replacement rules of January 2021 withdrawn.*
[19] *Sastry, N. S. K., & Arunachalam, T. (2025). Probabilistic safe RL: Modelling and managing risk in dynamic environments. EPRA International Journal of Research and Development, 10(9), 83-87.*

[20] Deloitte Touche Tohmatsu India. (2022). Mental Health and Well-being in the Workplace: A Survey of the Indian Workforce. Deloitte India, Mumbai.
[21] World Economic Forum & Harvard School of Public Health. (2014). Economics of Non-Communicable Diseases in India. World Economic Forum, Geneva.
[22] Employees' State Insurance Corporation. Sickness Benefit: rate, duration and contribution conditions. Ministry of Labour and Employment, Government of India.
[23] Ministry of Electronics and Information Technology, Government of India. The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023) and the Digital Personal Data Protection Rules, 2025, notified 14 November 2025.
[24] Ministry of Labour and Employment, Government of India. The Code on Wages 2019, the Industrial Relations Code 2020, the Code on Social Security 2020 and the Occupational Safety, Health and Working Conditions Code 2020, brought into force with effect from 21 November 2025.
[25] Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. Administrative Science Quarterly, 24(2), 285-308.
[26] Bakker, A. B., & Demerouti, E. (2017). Job demands-resources theory: Taking stock and looking forward. Journal of Occupational Health Psychology, 22(3), 273-285.
[27] Johns, G. (2010). Presenteeism in the workplace: A review and research agenda. Journal of Organizational Behavior, 31(4), 519-542.
[28] Goetzel, R. Z., Long, S. R., Ozminkowski, R. J., Hawkins, K., Wang, S., & Lynch, W. (2004). Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers. Journal of Occupational and Environmental Medicine, 46(4), 398-412.
[29] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., and others. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.
[30] International Labour Organization. World Employment and Social Outlook: Trends. International Labour Office, Geneva.
[31] Deloitte. (2022). Mental Health and Employers: The Case for Investment - Pandemic and Beyond. Deloitte LLP, London.