Canadian Journal of Marketing Research
ISSN: 0829-4836
https://canadian-jmr.com/index.php/cjmr

Research Article

Employee Wellness Programmes in Manufacturing: A Critical Review of Causal Evidence, Productivity Outcomes and Job Satisfaction

Suraj Shivaji Dhajal[1*], Anil Pruthiraj Dongre[2] and Rajesh B. Jawlekar[3]
[1]Research Scholar, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India
[2]Professor & Dean, Faculty of Commerce & Management, School of Management Studies, North Maharashtra University, Jalgaon, India
[3]Director, Innovation Incubation & Linkages, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India

***Corresponding Author**
Suraj Shivaji Dhajal
(dhajalsuraj@gmail.com)

Article History
Received: 12s.10.2026
Accepted: 02.11.2026
Published: 15.11.2026

DOI: 10.61336/cjmr.1604.39

Abstract: Wellness programs for employees are becoming essential in manufacturing since occupational health, work organization, physical demands and psychosocial conditions can affect worker wellness and functional performance. This paper reviews the impact of programmes which enhance health in the workplace on productivity and job satisfaction in a manufacturing setting. A systematic critical assessment was conducted of peer-reviewed intervention studies, systematic reviews and authoritative occupational-health advice related specifically to Covid-19 pandemic and literature from 2020-2026 including relevant earlier studies. The existing body of research deals with both the promotion of physical health and risk management and stress management approaches. Programs that improve health behaviours, perceived wellbeing, work ability and other aspects of employee experience can be effective, according to the research. Nonetheless, the proof pertaining to absenteeism, objectively assessed output and job satisfaction consistently shows improvement. The differences in study design, participant selection, duration of programme, measurement of outcome and context of industry prohibits direct comparisons and causality interpretation. The findings revealed that integrating individual support with risk mitigation can enhance the efficacy of wellness programs for manufacturing employees. In addition, supportive supervision, employee involvement and execution consistent with shift schedules and production requirements are also favorable. The research also reveals that self-reported productivity as a proxy for verified production cannot address limitation and there are not enough rigorous longitudinal studies in manufacturing. Future studies should combine established job-satisfaction measurement instruments with objectively recorded measurement of output, quality, absence and safety indicators, while also exploring programme costs, equity of access and longer-term sustainability.

Keywords: Employee Wellness, Manufacturing, Employee Productivity, Job Satisfaction, Occupational Health, Presenteeism, Total Worker Health

Copyright @ 2026: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

INTRODUCTION

Manufacturing is a distinct environment in which human performance depends on the combined effects of physical tasks, machine-paced work, process variability, shift arrangements, production targets and the organization of safety. The relevance of employee wellness consequently extends beyond optional fitness sessions or health-awareness events. It encompasses ergonomic design, psychosocial risk prevention, access to occupational services, fatigue management and worker participation [1]. The World Health Organization (WHO) identifies work organization and psychosocial conditions as intervention targets rather than framing mental health solely as an individual responsibility. Its 2022 guidance distinguishes organizational interventions, manager and worker training, individual support and return-to-work measures. Similarly, the International Organization for Standardization (ISO) 45003 standard addresses psychosocial risk within occupational health and safety management systems [2]. In a factory, a programme that asks operators to exercise while leaving uncontrolled repetitive motion, excessive overtime or high noise exposure

unchanged has an incomplete causal pathway to sustainable wellbeing. The United States National Institute for Occupational Safety and Health (NIOSH) Total Worker Health framework accordingly combines protection from occupational hazards with broader promotion of health [3] . These sources establish a prevention-first approach, but they do not in themselves provide proof that an intervention increases finished-unit output or satisfaction scores in a particular facility.

The use of several non-equivalent performance constructs complicates the evaluation problem. Manufacturing productivity can be defined in myriad ways, including measure of labour-hours per conforming unit, units per paid hour, output per operating shift, OEE and value-added per worker. Productivity, capacity to work and presenteeism reported by employees are indeed valid measures of health-related performance, but cannot be treated as a product throughput surrogate unless shown so in a plant [4]. Job satisfaction is yet another outcome. It captures an employee's evaluative attitude towards work. It can respond to supervisory fairness, autonomy, pay, security, voice and working conditions with medical indicators remain stable. If bottlenecks are driven by equipment, it may be that an intervention improves perceived wellbeing without affecting objective line output. Conversely, a speed-up may cause output to increase, yet the workers job satisfaction drops [5]. The systematic review undertaken by Kekkonen *et al*. in 2025 found only five of the 33 work-ability interventions effective for productivity. Their analysis characterized productivity losses predominantly by absence from work and presenteeism as opposed to direct output in the manufacturing sector; hence the finding demonstrates useful evidentiary caution, not a numerical estimate of factory productivity [6].

Recent research offers contrasting empirical signals. Song and Baicker's 2019 cluster-randomized evaluation of a workplace wellness programme in a large retail workforce detected improvements in certain health behaviours but no significant differences across measured clinical outcomes, healthcare expenditure and major employment outcomes after 18 months [7]. Jones, Molitor and Reif's randomized Illinois Workplace Wellness Study likewise identified participation selection and little evidence of gains in many downstream employment and healthcare outcomes. By contrast, Marin-Farrona and colleagues' 2023 systematic review reported improvements in work ability and selected physical-health outcomes across physical-activity interventions, although interventions differed considerably in format, follow-up and target population [8,9]. Støren and Grønningsæter's 2024 review-of-reviews found positive presenteeism or productivity effects for 22 of 45 heterogeneous health-promotion interventions. None of these results can be transposed mechanically to assembly, foundry, fabrication, chemical-processing or continuous-process manufacturing; actual shift coverage, production constraints, exposure levels and employment arrangements are potential effect modifiers [10].

There is a need for a manufacturing-centred synthesis because evidence of general wellness is frequently discussed at the level of programme presence, enrolment or healthcare behaviour, as opposed to distinguishing between uptake of the intervention and any workplace benefits which can be attributed to this causally [11]. The design of the intervention and the target population counts in a mental-health synthesis too. Miguel *et al*., assessed meta-analyses of universal, selective and indicated workplace interventions and Rugulies *et al*. distinguished work-related determinants from treatment of individual symptoms. Based on these studies and ISO and NIOSH recommendations, wellness is interpreted as part of the work system rather than individual choice [12]. This paper uses a critical, traceable evidence map, examines which indicators should be suitable to substantiate productivity and satisfaction claims, describes the interventions and emerging digital analytics and identifies the research design for a robust plant-level evaluation. According to systematic-review evidence a comprehensive, reproducible database export, duplicate-removal files and blinding facilities must be represented to substantiate an exhaustive claim [13].

Aim

To critically evaluate the effectiveness and evidentiary strength of employee wellness programmes for enhancing manufacturing employee productivity and job satisfaction.

Objectives

- Classify wellness programme components
- Compare causal and observational evidence
- Assess outcome validity and manufacturing applicability
- Examine analytical and digital delivery methods
- Propose a rigorous implementation and evaluation agenda

Research Questions

- **RQ1:** Which interventions show credible effects on work-related outcomes
- **RQ2:** How should satisfaction and production efficiency be measured separately

- **RQ3:** What contextual conditions moderate success
- **RQ4:** Which design and reporting gaps prevent reliable generalization

Scope and Organization

English-language workplace intervention research, principally 2020–2026, plus foundational randomized studies and occupational-health guidance; Sections 2–6 present methods, five thematic evidence areas, interpretation, limitations and conclusions.

MATERIALS AND METHODS

Review Design

This work is a systematic critical evidence review and not a registered systematic review or a meta-analysis. It brings targeted studies about verifiable research interventions into line with primary experimental comparisons and professional practice advice. The synthesis uses well-defined procedures and thematic comparison but does not claim total database coverage. In ethical terms, it is unreasonable to recreate or fabricate numerical database identifiers and exclusions in the absence of database subscriptions, timely queries and a screening electronic log. In the synthesis, data from the study and data sources are used and the overlap of trials that comprise the review has been accounted for.

Data Sources

The publicly available publisher sites, PubMed/PMC records, DOI-linked journal sites, WHO publications, ISO standards catalogues and NIOSH institutional guidance provide the bibliographic and institutional verification points for the directly cited evidence. It was not carried out nor claimed an institutional subscription-level export or complete cross-platform screening of Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis, Google scholar and PubMed for a reproducible future database-wide search. When available, publisher or bibliographic metadata was used to check sources. It is important to differentiate searchable platforms from databases that are actually searched comprehensively.

Search Strategy

A targeted discovery of sources published on or before 16 October 2026, with focus on 2020–2026 whilst retaining landmark experimental studies published in 2019. Example strings for later fully systematic version are: ("employee wellness" OR "workplace health promotion" OR "Total Worker Health") AND (manufactur* OR factor* OR industrial OR production); ("workplace intervention" OR "ergonomic intervention") AND ("job satisfaction" OR "work ability" OR presenteeism OR absenteeism); ("digital mental health" OR "wearable" OR "predictive analytics") AND (employee* OR worker*) AND (productivity OR wellbeing); and ("occupational health" AND (randomized OR trial OR "systematic review"). For future indexed searches, limit to years 2020 to 2026, English, article/review/conference study, public health, management, occupational medicine, ergonomics and industrial engineering, then look for 2019 fundamental trial separately. The strings are documented as an executable protocol and not wrongfully labeled as platform-specific historical exports.

Inclusion Criteria

The evidence that was directly included came from a source that could be traced to a scholar or authoritative institution. Furthermore, it addressed some aspect of employee health, work ability, productivity, satisfaction, psychosocial risk, or the evaluation of an intervention. Finally, the evidence supplied adequate methods or scope to permit sufficiently cautious interpretation. Reviews and comparative studies with peer-review were privileged for intervention effects; international standards and public-health guidance were exclusively used for policy and conceptual grounding. Older main trials were allowed where necessary to interpret outcome claims.

Exclusion Criteria

Evidence claims did not include marketing or vendor success stories, uncited percentage improvements, opinions pieces without methods or data that could be analyzed, bibliographic claims that were not accessible or able to be verified, studies that had no relevance to the workplace and reports that only measured wellness programme uptake but were being put forward as causal evidence of manufacturing output. No unpublished exclusions receive numeric total assignment.

Study Selection Process

Search for pertinent peer-reviewed studies was purposely conducted. Each reference was checked for issues related to bibliographic metadata enabled by its DOI. Reference checking also involved assessing reputable occupational-health

publications. A total of fifteen distinct sources comprises the first-order research evidence base with traceable links, which is eleven peer-reviewed research publications and four institutional guidelines or frameworks. Those sources have been assessed for relevance with respect to manufacturing conditions, study design, endpoints & limitations. This review makes no claim to complete cross-database retrievals, duplication screenings or blinded eligibility assessments and thus, no unverifiable identification or exclusion counts are reported. The end product maintains the difference between original teachings, systematic reviews and institutional guidance and avoids counting studies within a systematic review as separate contributions to the core reference set.

Data Extraction and Classification

Data extraction was done for each publication for the author-year, intervention or recommendation, target population, research design, an industry scale when directly reported, endpoint definition, effect direction, evidence limitations and manufacturing transferability. The methods fell under different categories:

- Occupational systems and risk control
- Physical activity and work ability
- Psychological interventions and job attitudes
- Economic or productivity evaluation
- Digital and analytic delivery

Publications that synthesized different trial results contributed only one review-level total to the database to prevent double counting. Where missing, software platforms and statistical accuracy indicators weren't imputed.

Outcome Definitions and Evidence-Mapping Rules

An intervention effect should only be called a productivity improvement when the dependent variable has been specified before observing the outcome and the comparison addresses a counterfactual. For discrete manufacturing, a candidate measure is accepted units produced per direct paid labour-hour; an alternative is standard earned hours divided by actual hours, provided standard times are stable and contemporaneous product-mix changes are considered. Continuous plants may require tonnes of conforming material per staffed production hour, normalized for feedstock quality, planned outages and process yield [14]. Overall equipment effectiveness combines availability, performance and quality, but it is not by itself a pure employee-productivity index because mechanical availability and process capability can dominate its variation. A study claiming an employee wellness effect from OEE alone needs an explicit theory identifying which components employees can reasonably influence and must account for concurrent maintenance and automation programs. Human-resource outcomes such as sick leave, early exit, willingness to recommend the employer and self-rated performance have value but should be labelled separately. These definitions are analytic recommendations derived from the general distinction between employee health, work ability and productivity made in prior reviews [15].

Job satisfaction should be specified as a distinct psychosocial construct rather than inferred from program satisfaction. An employee who appreciates free health screening may remain dissatisfied with overtime distribution, job security or supervisor behaviour. Validated multi-item instruments are preferable to an unaudited single question because they permit assessment of internal consistency and change in clearly defined facets [16]. A manufacturing field trial should report the scale, response range, scoring orientation, language adaptation, missing responses, test timing and pre-specified minimally important differences if established for that instrument and population. When a scale has not been validated for a particular language or group, researchers should describe content review, pilot testing and factor analysis as necessary; a high Cronbach coefficient alone does not establish construct validity. Aggregating responses to a department mean should also be justified by the structure of the measure and supported by intraclass correlation where practical. These safeguards are especially important when permanent and contract labour differ in access to benefits, rights and perceived safety [17].

Wellbeing measures might include musculoskeletal pain, validated psychological distress, perceived work ability, recovery, fatigue or sleep. Their implications for the ultimate outcomes differ regarding their causal position. For instance, fewer symptoms may increase work capacity [18]. However, we also see that actual output is dependent on demand and production-system constraints. When multiple intermediate and final endpoints are evaluated, selective emphasis on the most favorable one can lead to misleading conclusions. The protocol should therefore differentiate between primary endpoints and secondary outcomes and exploratory mediators; apply appropriate adjustment for multiplicity, or explain why endpoints are interpreted descriptively; show both point estimates and interval estimates [19]. The absence of statistical significance does not equate to evidence of no

effect, especially with short follow-up, small clusters or variable uptake, it is important to look at the uncertainty bounds and the plausible effect sizes. These considerations can help explain why the work-ability and wellness reviews observed different patterns of positive outcomes [20].

Suggested Data Collection Protocol

A plant-level research protocol should start off with a workplace exposure and implementation audit. At baseline, teams should record the task repetition at the workstation, manual load, environmental hazard, break time, shift length, work-rest cycle, changeover, practices of supervisor and whether person has access to service or on-site health services during paid time [21]. The logic model for intervention should link each identified need to a tangible control, intermediate outcome and final outcome. Redesigning equipment can help to focus on the duration of awkward postures, muscles ache, absence and eventual throughput, while a supervisor training programme may concentrate on psychological safety, employee voice and satisfaction [22]. It is necessary to note the number of workers invited, eligible, enrolled, exposed to the intended dose and followed-up. Further reasons for non-participation shall be recorded if voluntarily supplied. The information distinguishes programme feasibility from treatment effectiveness and is consistent with participatory and prevention-centred occupational principles.

In a practical stepped rollout, matching arms, or sites, can be initiated in a control condition and switch over to the intervention at after randomized times. A difference-in-differences estimate can be considered when genuine randomization is not possible, but the parallel-trends assumption must be supported by a number of preintervention periods, not merely asserted. Measures of production should be sourced from verified manufacturing execution or enterprise records, audited against timekeeping and quality records. The analysis should reveal holidays, temporary shutdowns, seasonality, product-mix changes and simultaneous workforce restructuring. To allow for inside-person surveys without compromising clinical confidentiality, your anonymous identification keys and appropriately restricted linkage are used in all surveys. Caution is needed to ensure that disaggregated outcomes do not compromise the privacy of small night-shift or contractor cells. Plant managers should look at the safe aggregates, not a health-risk score for an individual unless there are unusual legal and ethical requirements that clearly justify a different design.

Three assessments (baseline, process and impact) allow for interim impact learning that misinterpretation of signals may avoid. It's a design advice; it is not fixed evidence minimum. The calculation of sample size should take into account the allocation unit, which has cluster correlation, expected turnover and missing surveys; It would be overly optimistic to treat hundreds of operators within a small number of randomised production lines as being statistically independent because that will give false precision [7]. The intention-to-treat analysis will retain the effect of the offer of the programme. In addition, further analyses based on adherence may assess mediation mechanisms with suitable selection caveats. Changes to the production denominator and the exclusion of major shutdown periods must be reported in the sensitivity analysis. When programme uptake is low, the offer effect and the measured fidelity both matters. The former answers the policy question, while the later diagnoses implementation failure.

Statistical and Economic Analysis Framework

A proper main analysis might employ multilevel regression with worker and line-level features along with either clustered standard errors, or a mixed-effects structure aligned to the randomisation and outcome distribution. Data on sick leave may require negative-binomial modelling if overdispersion is apparent; proportions of defective output shouldn't be treated with ordinary least squares uncritically but modelled via binomial or appropriately weighted modelling. After checking the scale properties and baseline imbalance, we can use a suitable continuous/ordinal model to compare work satisfaction composites. Analysis report conclusions must include details about the selected estimand, effect size, various levels of confidence intervals, as well as loss and missing data strategy. Writers should refrain from coming to conclusions about differences based on subgroup interactions resulting from cells that are too small and hence have inadequate power. Journal authors estimating causal mediation between health changes and productivity must indicate first that the relationship between mediator and outcome may be corrupted despite the fact that the effects have been randomized.

Cost-consequence analysis is more transparent than forcing everything into one Return on Investment (ROI) figure. Cost of intervention, hours of participation paid for, occupational health staffing, coverage of replacement shifts; whether there has been gain or loss in actual labour hours, effects of turnover and whatever you can think of can be shown in parallel if there is one clear net benefit monetized. If we are trying to calculate net benefit in monetary terms labour must not be counted for one period of absence many times and economic gain due to sales or equipment changes must not be allocated to their degree of influence. Analysts ought to specify the currency year, point of view and everything they are going to take into account. In cases where clinical or claims data are out of reach this should be stated rather than making up figures by relying on savings related to totally different

fields. Marin-Farrona *et al*. and Kekkonen *et al*. both point out the gaps in existing economic evidence and therefore it is necessary to follow the recommendations mentioned above [6,9].

Digital tracking adds another statistical layer. A model that predicts sickness absence from past absence, shift schedules and self-reports may be evaluated with discrimination, calibration, positive predictive value and subgroup error; however, the performance must be tested on held-out sites or temporally separate observations. Testing whether the prediction system improves an outcome requires random assignment of management action based on model alerts. It is best to exercise caution in interpreting accuracy when adverse outcomes are rare. A model that predicts no cases can have a high accuracy and fail to identify any worker who might benefit from additional support. Consequently, the f1-score, recall, calibration and decision-curve usefulness should be selected to suit the decision and class balance. The fact that the model predicts an endpoint of operational interest does not justify the repurposing of personal health information for employment discipline.

Transferability and Reporting Standards

One should evaluate the external validity instead of assuming it. The trial of an university or a retail employee is quite different in baseline movement, exposure profile, autonomy, ability to interrupt tasks and supervision compared to metal fabrication, food processing or electronics assembly. Rifts also emerge between automated process-control suites, manual assembly teams, logistics crews and maintenance workers within one mill. Researchers must mention only ethically sound aggregates disclosing sector, region, size, contract, shift, hazards, baseline physical demands and demographics of workers engaged in the study. A program that modifies perceived work ability in a low-load office population may reduce manual handling injuries but we don't know; thus, targeted ergonomic changes may be very relevant to that physical task but have no reason to improve morale in that other low-exposed group.

The programme name is as important as implementation description. The complete intervention requires the identification of the responsible professionals; the frequency, duration, paid/unpaid participation; location, time allocation, supervision and modifications to the working environment; and whether co-design with workers happened. The comparators ought to detail what would normally happen through occupational healthcare and the existing safety controls. This is so that it is clear what the added content of the intervention is. Research should report how shifts, crossovers and contamination across lines may have affected engagement and withdrawal, as well as any unintended adverse effects like social pressure to disclose conditions, inequitable access or work intensification. These particulars allow for reproduction and mark a context-sensitive prevention mechanism as distinct from a non-specific wellness marketing package.

Reporting has to adhere to the principle that there are three separate claims: the intervention was implemented; it improved a chosen measure of worker wellbeing; and it causally improved labour output or satisfaction. The first requisite calls for processed data. The second necessitates valid individual outcomes. The third stipulates outcome-specific causal evidence. Neither following the regimen nor a marked change in self-reported exercise acts as a substitute for the third claim. If we clearly identify these claims, many of the ambiguities shown by randomised evaluations and systematic reviews will go away. Market-valued occupational-health interventions will still be treated as intrinsically valuable even without short-run productivity change [7].

For synthesis specific to manufacturing, a study should be separately coded for four axes:

Intervention Level: Individual behaviour, workstation, team scheduling or organisational policy
Outcome Level: Symptoms, work ability, subjective productivity, objective operational performance or satisfaction;
Design Credibility: Randomisation, concurrent comparator, longitudinal adjustment or uncontrolled description
Implementation participation, shift coverage, adherence, supervisory support and concurrent process changes.

Never take a positive finding on one axis as success in the others. According to the differing experimental and industrial evidence, this framework is proposed as an analytical classification and not by a proven predictive model [23].

Related Work

Conventional Wellness, Health Promotion and Occupational Risk Control: Traditional employee wellness strategies typically focus on individual behaviours with health screening, education and exercise nutrition counselling and being voluntary. Song and Baicker [7], posed a strong challenge to a simple 'wellness causes productivity' narrative. Their cluster-randomised programme in a large retailer improved selected health behaviours, but did not establish a corresponding improvement in measured health expenditure or various employment outcomes in its follow-up [7]. Jones, Molitor and Reif [8] also randomized access to a full employee wellness program in a large university and showed employee selection into participation is a material issue. Employees who choose to participate in health programs may already differ in health

or tastes from non-participants. Both studies are useful precisely because random assignment reduces the confusion inherent in cross-sectional satisfaction surveys; neither was done as a production line experiment [8].

The standard view of a factory must begin with avoiding avoidable hazards. The guidelines suggested by WHO in 2022 deals with work organisation and psychosocial risk The ISO 45003:2021 uses a safety-management system to locate psychological health while NIOSH Total Worker Health explicitly focuses on eliminating unsafe working conditions while promoting broader wellbeing. There are realistic avenues to increase satisfaction through improved perceived protection, supervisor credibility and worker voice. Nonetheless, an operational standard isn't an effectiveness trial: the adoption or certification does not demonstrate a statistically significant change in the output of the unit [24,25].

The gap in evidence identified in these studies is crucial: the justification for the implementation of production programs is based on evidence accumulated in unrelated industries, while working practices, risk exposure and motivation are not measured simultaneously [26]. In order for the control group to be valid, real-life risk mitigation techniques, operating conditions and equipment condition should have been taken into account in each production site instead of comparing the wellness group with the control group as if the selection of participants was random [27].

Physical Activity, Ergonomics and Work-Ability Programmes

According to Marin-Farrona *et al*. [9], 860 records obtained from electronic search were potentially relevant and 16 randomised controlled trial papers evaluated workplace physical-activity programmes The report showed positive outcomes for work ability along with chosen physical health outcomes particularly fitness and musculoskeletal indicators however variance in modality, workforce, programme duration and instruments far from "dominant activity protocol" identification. Due to the lack of reporting of cost and other economic information, their synthesis could not effectively resolve cost-effectiveness. A factory application should distinguish between genuine lifting tolerance or pain developments and verified increases in acceptable units per labour-hour; production rates may be set by tooling or takt time rather than aerobic capability [9].

Støren and Grønningsæter [10], analyzed broader worksite health promotion evidence from 2,196 records, keeping 14 reviews. Out of 45 different types of workplace programs, 22 were categorized by the authors as having positive presenteeism/productivity outcomes. More specifically, favourable evidence substantially involved multicomponent approaches. The study also noted that presenteeism definitions vary and that targeting the individual's lifestyle is frequently preferred to work-organisation change. These are second-order aggregate findings and do not imply that a single treatment effect operates across factories; the programmes which feature a physical activity, social support and ergonomic change differ both mechanistically and by measurement of effects [10].

Kekkonen and his colleagues [6], evaluated the effectiveness of different work ability interventions. After eliminating 26 of the 55 articles that were analyzed for quality, the authors got the collection of 29 papers concerned with 33 studies of work ability interventions, where only five of them had statistically significant results and the rest had no effect. The review was designed without taking ergonomics interventions into account and therefore it cannot disprove the usefulness of ergonomics. Comparing the three reviews reveals the denominator issue because the effectiveness of the reviews was determined by different criteria. The research deficit in this sphere exists because of the lack of controlled investigation of the workers' body load connected with their occupational exposure, satisfaction and production quality verification [6].

The numerical denominators refer to different units' records, reviews, intervention categories or evaluated interventions and must never be combined into a single overall success rate (Table 1).

Mental Health, Organisational Design and Advanced Digital Approaches

The authors of the current report led by Miguel *et al*. [12], synthesized 16 meta-analyses of universal, selective and indicated workplace mental-health interventions, including 66 pooled effects. The results were focused on symptoms and positive mental health, while work-related coverage was less extensive. The evidence supports making psychological assistance readily accessible to all employees; however, this alone cannot rank an employee-assistance counselling programme against a task redesign as a manufacturing productivity intervention [12].

The paper by Rugulies and colleagues [28], examined some work-related causes of mental-health conditions and strategies for the workplace to improve them, such as high demands, low control, adverse social interactions and organizational prevention. According to, the ISO 45003 standard considers psychosocial hazards to be workplace exposures that can be managed [28].

In terms of actual technology innovations, some digital self-assessments, discrete mental-health aids, fatigue-wearable indicators, electronic incident dash boards and predictive modelling of absence risk. Workers rotate through shifts and have no easy access to a daytime clinic; Such tools might help. Still, a classifier by artificial intelligence could be accurate in predicting self-reported stress but not help improve job satisfaction or output: classification accuracy, F1

score and area under the curve describe prediction, not intervention benefit [29]. The current set of verified manufacturing trials does not show that machine-learning wellness prediction results in a measurable productivity lift. The right entry for the ML accuracy, the error metrics and the plant-level cost savings is "Not reported" and a credible-looking manufactured percentage [30].

The model-development gap includes properly validated consent and privacy protocols, independent partitions of data across workers and facilities, evidence of calibrated prediction and randomized tests of the decisions that analytics enable. Monitoring should not be individual surveillance which is punitive as it may harm safety reporting, programme trust and satisfaction. An effective androidx architecture focuses on aggregated anonymous trends and risk mitigation, with different access to individual clinical details [31].

Pehlevan and Şevgin [32], conducted a randomized controlled trial among employees with low-back pain in a plastic-moulding factory. Both groups received ergonomic training; the intervention group also performed stretching and core-stabilization exercises three times weekly for eight weeks [32]. Between-group comparisons favoured the combined programme for pain, disability, fatigue and sleep-quality instruments, including the McGill Pain Questionnaire, Oswestry Disability Index, Fatigue Severity Scale and Pittsburgh Sleep Quality Index. Crucially, factory output, defect rates and overall job satisfaction were not established as outcomes, so an improvement in symptoms should not be re-labelled as a verified productivity effect [33].

Lerche and colleagues in 2024 tested another mechanism: changing how physical activity is distributed during work, instead of adding a session of exercise. In a cluster-randomized trial conducted over a period of 12 weeks, 83 workers from a total of 28 work teams were allocated either to a Goldilocks Work intervention or to resume usual work. The allocation was intended to be 60% sitting, 30% standing and 10% other active tasks doing tasks every hour. Although there was a reported delivery fidelity, sensor-assessed physical behaviours did not change as planned and there was no statistically significant effect on productivity (estimated contrast −0.03; 95% confidence interval −0.77 to 0.72). The lack of results questions a hypothesis that a theoretically achievable balanced mix of tasks is sufficient, without feasible changes to production planning [34].

Jin and colleagues [35], conducted a study that involved 181 employees who worked at a rail-vehicle assembly plant. The study ran for one year after which the researchers implemented ergonomic instruction, auxiliary tool improvement and also the distribution of ergonomic education materials. Scores for back, neck, shoulders/arms, wrists/hands on quick exposure check decreased ($p<0.01$) [35]. Reported exposure to awkward postures and musculoskeletal symptoms also decreased. The long follow-up and industrial specificity strengthen practical relevance. Nonetheless, in the absence of a randomized comparator or directly measured throughput and job-satisfaction endpoints, secular changes, co-interventions and reporting effects remain plausible alternative explanations. As a package, these three industry studies show that intervention mechanism, implementation fidelity and endpoint selection together determine the strength of claims [36].

Table 1: Summary of Selected Intervention Evidence

Study	Method/model	Population / scale	Outcome measures	Verified finding	Limitation
[7]	Cluster-randomized wellness trial	Large retail employer not factory	Health behaviour, clinical and employment outcomes	Selected behaviour gains; no significant gains across several other outcomes	Sector and follow-up specificity
[8]	Randomized wellness access	University employees	Medical spending, work outcomes, participation	Limited downstream effects; selection important	Not production operations
[9]	Systematic review of RCTs	860 retrieved; 16 included papers	Work ability, fitness, productivity	Selected health and work-ability benefits	Modality and cost heterogeneity
[10]	Review of reviews	2,196 records; 14 reviews; 45 programme categories	Presenteeism/productivity	22 of 45 programme categories favourable	Non-uniform definitions; overlapping source studies
[6]	Systematic review	29 articles; 33 interventions	Absenteeism and presenteeism	5/33 significant intervention results	Ergonomics excluded; varied populations
[12]	Umbrella review of meta-analyses	16 meta-analyses	Mental-health and some work outcomes	Benefits vary with target and endpoint	Work outcomes less consistently reported
[28]	Narrative evidence review	Cross-sector occupational evidence	Work-related mental-health risk	Emphasizes organizational determinants	Not a factory-specific effect estimate
[32]	Factory randomized trial; ergonomics vs ergonomics plus exercise	Plastic moulding workers; 8 weeks	Pain, disability, fatigue, sleep	Combined treatment significantly better on reported scales; exact effect size not reported here	No verified output or job-satisfaction effect
[34]	Cluster-randomized work redesign	83 industrial workers; 28 teams; 12 weeks	Pain, sensor behaviour, productivity	Productivity contrast −0.03 (95% CI −0.77 to 0.72); no significant effect	Intended task mix not achieved
[35]	One-year ergonomic pre/post intervention	181 vehicle-assembly workers	QEC, postural exposure, musculoskeletal symptoms	Regional QEC scores declined ($p < 0.01$)	No randomized comparison or verified output

Comparative Performance, Evaluation Measures and Causal Inference

Success of interventions is highly sensitive to what is measured. Song and Baicker [7], directly estimated intervention effects based on randomised assignment; Jones *et al*. [8], examined downstream economic and employment variables, revealing participant-selection problems. The studies conducted by Marin-Farrona *et al*. [9] and Støren and Grønningsæter [10], synthesized diverse work-ability and presenteeism measures. Kekkonen *et al*. [6], used a restrictive risk-of-bias filter, which resulted in a very much smaller set of positive impacts on productivity. According to the aforementioned reference, this does not imply older supporting results were erroneous; rather, a health benefit, a positive within-group shift and a controlled productivity effect are distinct claims [7,9,10].

We must pay special attention to production-based indicators. Labour productivity ideally uses conforming output divided by paid direct labour-hours, with rework, scrap, overtime and plant use modelled. Days absent from work, turnover, quality defects and "near misses" in safety should be correlated and recognized with proven job-satisfaction and perceived-support tools. Relevant covariance can include things such as production mix, capital investment, order volume maintenance shutdowns, seasonality, work team, age distribution and baseline workload [37]. A staggered stepped wedge evaluation or cluster randomization can minimize contamination and allow the treatment to be isolated from shocks to demand or equipment when the intervention is rolled out by production line or site. The rationale for these choices comes from the studies reviewed, although no comparative effect size is claimed for these indicators [38].

Cost claims are especially fragile because wellness programme costs include worker paid time, specialist fees, shift backfill, implementation and equipment. A defensible business case should report incremental net benefit and uncertainty under competing productivity valuations. Reviews indicating missing cost data must not be rewritten as evidence of positive return on investment. The main gap is harmonized reporting of health, satisfaction, safety, hard output, implementation fidelity and cost in the same manufacturing sample [39] (Table 2).

Distinct evidence levels and endpoints preclude ranking these approaches by a common percent change. The most persuasive evaluation couple's random allocation with predefined factory output and satisfaction endpoints.

Table 2: Comparison of Methods, Datasets and Evaluation Parameters

Approach	Study evidence	Observed/assessed variables	Reported quantitative result	Factory transferability
Randomised general wellness	Song and Baicker	Behaviours, health outcomes, costs, employment	No general productivity gain established	Moderate methodological relevance; low sector match
Randomised wellness choice	Jones *et al*	Selection, utilisation, work outcomes	Manufacturing output: Not reported	Useful for enrolment bias control
Exercise RCT synthesis	Marin-Farrona *et al*.	Work ability, fitness, musculoskeletal outcomes	16 trials; production quality: Not reported	Health pathway plausible; direct output uncertain
Presenteeism review synthesis	Støren and Grønningsæter	Self-rated productivity/presenteeism	22/45 programme classes positive	Measurement heterogeneity high
Work-ability effectiveness review	Kekkonen *et al*.	Absence, presenteeism and costs	5/33 significant; 28/33 non-significant	Strong caution about effect inference
Mental-health meta-analysis umbrella	Miguel *et al*	Symptoms, wellbeing, selected work metrics	16 meta-analyses; 66 pooled effects	No common factory throughput endpoint
ML/wearable programme	No qualifying comparative source in core corpus	Accuracy, F1, productivity, satisfaction	Not reported	Requires trial validation
Exercise plus ergonomics	Factory low-back pain randomised trial	Pain, disability, fatigue, sleep	Significant relative improvement; exact estimate not extracted	Health benefit does not equal output gain
Goldilocks task redesign	Industrial team cluster trial	Wearable behaviour; productivity	−0.03 (95% CI −0.77 to 0.72) for productivity	Implementation failed to shift target activity mix
Factory ergonomic package	Vehicle assembly longitudinal cohort	Quick Exposure Check; symptom questionnaires	181 employees; QEC $p < 0.01$ by major region	Objective production effect not assessed
Three-year wellness trial	Multisite cluster randomization, non-manufacturing	Health behaviours, absenteeism, job performance	No detectable employment outcome differences	Transfer to production needs testing

According to Song and Baicker [36], they revisited their cluster-randomized multisite wellness intervention at three years which provides a longer-term counterpoint. The Nutrition, Movement and Stress reduction program (NMCS) were associated with statistically significant improvement in selected self-reported health behaviours, but there were no detectable differences in clinical indicators, health care use and spending, absenteeism, tenure and job performance. Although the employer was not a factory, the causal leverage of this result is greater than an uncontrolled pre/postfactory survey. The inference does not imply that wellness is ineffective; it means that the impacts of intervention are outcome-specific and not automatically transferrable across domains or over time [7].

It is useful to distinguish between precision and significance in interpretation. The parametric test that estimates the coefficients with confidence interval for example the estimating sign and its certainty. Similarly, change on a symptom scale that is statistically significant is not operationally important. Manufacturers must state the minimally important difference for health scales and report the economic value of one unit of conforming output while bearing in mind staffing and machine availability. Composite scores can hide a negative trade-off between wellbeing and production outcomes without alignment [40].

Manufacturing Implementation Challenges, Equity and Emerging Research

Fidelity in implementation are likely the more consequential than branding of the programme. In the absence of relief operators, when maintenance windows are narrow, breaks are staggered, or when staffing is outsourced, production employees may not be able to attend. A programme available only to administrative staff working day shifts could increase – rather than reduce differences in occupational wellbeing. Participation must be encouraged and NIOSH has a structure while WHO will tell you to prevent from the organization. When involving our producers in the continuous production of a show, among other principles, we say night-shift delivery, translation where needed and fair access for contracted personnel [41].

Factors that affect job satisfaction often get left out. Even if pain has been credibly reduced, satisfaction may not increase if supervisors use the freed-up capacity to increase pace of work without involvement of workers. On the other hand, it is possible that improved satisfaction may occur because of the more predictable shifts experienced and participatory control, even though line throughput remains equipment-constrained. The outcomes reported in the cross-sector intervention literature do vary, but that does not make them inconsistent. The pathways and effect moderators at play differ. ISO 45003 provides a governance vocabulary for psychosocial risk. Moreover, workplace experimental studies highlight that programme presence alone does not equate to effect evidence [42].

Given these considerations, it is recommended that future research agenda favour preregistered multisite evaluations, triangulated individual and machine-level indicators, disaggregated subgroup analysis and longer observation windows. Transparent decisions about data minimization, model fairness and independent monitoring enabled by digital delivery can expand access. Above all, it is important to first consider if a manufacturing programme decreases exposures and improves worker welfare, not whether the benefits can be attributed to production. A health workforce capable of sustaining performance safely is a better target than seeing health interventions as automatic output multipliers [43].

Each gap is an actionable design requirement, rather than a claim that any particular untested manufacturing solution has already demonstrated an effect.

For a vehicle assembly plant, the most convincing hypothesized pathway from health to productivity may concern occupational exposures that repetitively disrupt task performance. Discomfort which negatively affects work capacity can be caused by reaching, force, repetition, restricted recovery and fatigue. However, the unit of analysis will be determined by the production system. A pre-post comparison of total cars produced will be particularly weak if there were within period changes in model-mix, automation or supply interruptions. The outcome architecture should therefore separate line takt adherence, conforming assemblies per paid operator hour, rework, restricted duties, discomfort and satisfaction. In the event of an ergonomic redesign of workstations that differs systematically from comparisons, it is important to have matched baseline trajectories or allocation by work cell. A health effect should not be deemed failed just because of an upstream supply constraint. Similarly, a rise in final output without change should not be attributed to exercise classes [6].

Manufacturing plants for food, pharmaceuticals and other hygiene products impose constraints for protective clothing, sanitization, environmental conditions and continuous quality assurance. An intervention can involve things such as rest scheduling and safe break areas, hydration or redesigned rotation. However, time to clean equipment, batch sizes and yield loss can change independently. The endpoints of the study should include quality-adjusted output, reportable absence, job satisfaction, fatigue and safety. To protect the confidentiality of health responses, they should be kept out of the normal quality and disciplinary systems. The scheduling of the program should ensure that workers on the night shift get equal access and control staff do not unintentionally lose their existing legally mandated protections. When calculating compliance for a trial of shift-based delivery, eligible paid shift opportunities should be used, rather than total number of enrolled workers, as an equal number of available sessions may obscure unequal effective access. The operational interpretation translates the prevention-first orientation of WHO and NIOSH into performance measurement See WHO 2005 but does not claim these organizations have assessed a quantified effect for this manufacturing sector [44].

Plants that are heavy manufacturers and maintenance-intensive may experience a distinct combination of thermal exposure, awkward postures, lifting as well as noise and vibration. Moreover, there is unpredictable demand for repairs.

Mechanical lifting assistance or improved ventilation are examples of upstream controls. Conceptually, these are distinct from an individual coaching programme. Exposure measurements and injury-risk indicators as well as job attitudes should be used to evaluate possible benefits. To maintenance staff, comparable units of output completed standard work orders adjusted for work order complexity and unplanned breakdowns; counting work orders gives a reward for fragmentation into many small jobs Results on satisfaction may respond to hazard reduction and management responsiveness but cannot be inferred from fewer incident reports as the better reporting culture could improve the recorded near-miss counts. Quantitative evaluations should use balanced safety and quality outcomes so that less reporting of events is not mistaken for less risk. The Total Worker Health and ISO 45003 provide coherent prevention and psychosocial frameworks, but their formal effect sizes require testing in primary plants [45].

The participation mechanism is critical for women employees, older workers, migrants, PWDs and outsourced operators, as workplace conditions and access to schemes vary. An overall average across a site may conceal what is happening to subgroups in opposite directions. For example, it may be that an intervention works among the well-resourced day-shift operators, but that effective participation is practically impossible in a small night-shift unit. Instead of presupposing which group will gain from your intervention, researchers should document baseline barriers, invite meaningful participation from the workers, and, where possible, plan sufficiently powered subgroup comparisons. It is necessary for medical information to be private and kept away from supervisors. Researchers ought to implement aggregate reporting thresholds that safeguard confidentiality and should inquire of workers whether their involvement entails social pressure, lost pay, or conflict with production quotas. It is important to keep the distinction between eligibility, invitation, initial enrollment, attendance and sustained use visible. Insights from the randomized wellness evaluations suggest that self-selected attendees should not be treated as interchangeable with non-attendees – this is so even after observed demographic adjustment [46].

A competing interpretation regarding employee agency is important. Wellness initiatives offered in organizations as a reward or individual lifestyle projects are easy to promote and count. A manufacturing worker may instead experience one main source of distress: high demand, poor task discretion or insecure scheduling [28]. If these latter factors remain unchanged, a programme focused on the individual may lead to changes in knowledge or healthy habits without any change in satisfaction. Organizational improvements can alter the employee experience long before any resultant absenteeism or economic impact can be. It means to evaluate in stages: assessing the work exposures and fidelity; state the evaluation of validated health and attitudes; and assess the effects on production and economic endpoints. A pre-planned staged design is desirable; one shouldn't use it heuristic-wise to rationalized unfavorable primary results post-data access. Evidence on work-related mental-health determinants supports prioritizing organizational intervention targets, whereas trial reviews warn against treating each intermediate change as evidence of wider benefits [47].

Programmes that include some kind of wearable monitoring, wellness application or dashboard analytics should be assessed against an appropriate analogue alternative, such as scheduled face-to-face access. A technical novelty does not cause an endpoint. An app collecting regular self-reports and providing managers with detailed operational visualisation won't advance employee wellbeing if no one does anything with the findings or workers distrusts the app [48]. Assessment needs to capture the eligibility of devices, the accuracy of sensors in real working environment, their missingness under PPE, consent rates, false alerts, additional administrative burden and differences between shifts. When evaluating a model, if the workers whose responses are used to test share individuals with workers whose responses are used to train the model, the result may overstate the generalisation of the model. If individual advice is provided during deployment, the trial should also measure whether that advice would be practicable on shift. An effectiveness claim is meaningful only when it involves a measure of intervention allocation and downstream health or satisfaction. That is, it is not enough to be able to estimate the accurate fit of a predictive model. The trend is also noted throughout the core reviews, a limitation of inferring occupational outcomes from proxy measures [49].

Wellness programmes are multi-stakeholder initiatives that may benefit some more than others. Fatigue relief and a considerate management team may appeal to workers; reduced risk and sustained work ability may appeal to occupational practitioners; and sick leave and production cost may appeal to financial managers [50]. A proper evaluation presents these dimensions alongside one another rather than collapsing them into an arbitrary weighted score that can mask satisfaction deterioration with a large output gain. There are some consequences that ought to be considered conditions instead of trade-offs: the need for protective measures cannot be taken for granted. From this point of view, health results and employee voice should be assessed as meaningful benefits. At the same time, economic results should be presented in a straightforward way as far as possible. The evidence base for ethical and reproducible production will thus depend on the methods taken from industrial engineering, occupational medicine, ergonomics, behavioral sciences, trade unions and the independent movement. Existing reviews do not allow determining the possibility of any wellness program being appropriate for each factory because of the high cost of its use [39].

Table 3: Research Limitations and Identified Gaps

Domain	Current limitation	Risk to inference	Recommended study design
Sector transfer	Retail/university and mixed-sector studies dominate landmark evidence	Factory effect extrapolation	Multisite plant cluster trials
Outcome definition	Presenteeism, work ability and productivity conflated	Non-comparable effect directions	Prespecified output + satisfaction + health endpoints
Programme selection	Voluntary enrolment and self-selection	Healthy volunteer bias	Intention-to-treat, randomized offer
Follow-up	Different duration and attrition	Short-term effects overstated	At least several repeated time points; disclose retention
Economics	Missing implementation and indirect costs	Unsupported ROI	Pre-registered cost-consequence and sensitivity analysis
Technology	Prediction treated as intervention efficacy	Deployment optimism	External validation plus randomized decision-support evaluation
Inclusion	Shift, contract and language differences	Inequitable programme access	Shift/contract subgroup reporting
Data governance	Wearable and clinical data sensitivity	Trust and privacy harms	Consent, minimization, aggregate monitoring

Program evaluations are difficult to interpret because of the distinction between immediate responsiveness and durable change. Employees may initially give a thumbs up just because the management offers some sort of attention. Also, new facilities and new convenient services. Subsequent impact relies on the service's continuity, engaging staffing, ergonomic upkeep and supervisors' external evaluation [51]. A rational follow-up plan should therefore record both early responses and later retention of benefits, with comparable measurement frequency in associated departments. Researchers should be cautious about only presenting the most favorable month in their outcomes or effects if they diminish as shift pressures return. The same reasoning applies with the measurement of productivity (e.g. production ramp-up and learning curves after an intervention may first lead to depressed output even if the process eventually becomes safer or more sustainable) [52]. Without an extant time period and examination of adaptation, one might interpret that temporary implementation cost as failure. On the other hand, early reductions in sickness absence could be attributed to random month-to-month variation unless marked and compared to a suitable counterfactual. These considerations offer a reasonable rationale for the observed dissimilarities, without resorting to claiming unreported magnitudes of longitudinal effects

Displacement is another problem faced by us. An intervention may transfer workload from one category to another. In instances such as when wellness sessions are covered by relief operators or when selected employees receive lighter work while other employees carry heavier workloads. An employee participation system that focuses exclusively on those who take part can provide an exaggerated overall benefit and cloak a rising workload or discontent elsewhere. All impacted units should form the analytical population; whether an increase in attendance, throughput, or satisfaction in one group coincided with a change in another [52]. The treatment effect at the department level can capture more redistribution than individual outcomes among attendees, whenever it is operationally feasible. The entire workflow should also analyze safety indicators. If manual activities have only shifted downstream, a decrease of the symptom for one station does not mean lower ergonomic load. Total Worker Health concentrates on integrated organization-wide approaches focusing this issue on the design of operations systems [53].

Adaptation of the program at various plants creates a tension between the two - standardization and local fit. A rigid method of intervention allows comparisons to be easier; however, it might not be practical due to different shifts, hazard profiles or labour arrangements of the facilities. An adjustable intervention that protects essential functions – worker involvement, occupational risks controlling, confidential and regular feedback, probably could focus retrofitting session delivery, translation, scheduling and site-specific exposure controlling [54]. Evaluators should record what elements are core, which elements were changed, the reasons for the changes and was adherence adequate. A mixed-methods design can capture explanatory information from worker interviews, process observations and supervisor records and retain quantitative outcomes for effect estimation. The subjective results should not replace the objective effects, but they may explain why different programmes with similar inputs have different results [55]. For any pooled comparison of effectiveness to be defensible, the intervention descriptions must be made transparent.

When productivity valuations assume that every absence day avoided is a further day of output, they can distort the evidence further. Absorption of an absence through overtime, redeployment or automation may not imply costs identical to a naive wage-times-days calculation for some plants. In other cases, an absent specialist may constrain the entire bottleneck operation and cause bigger indirect disruption. Thus, an economic evaluation must indicate whether it uses wage-based human capital costs, actual replacement costs or a narrowly-friction based one and must avoid summing alternatives that represent the same lost work. Staffing continuity benefits may have a delayed impact, while service delivery cost impacts are immediate. The time frame of reporting and the allocation of costs among the employer, employee and health care system prevents an apparently favorable

business case to mask unreimbursed employee time. Systematic evidence already available specifically denotes gaps in cost reporting; thus no universal manufacturing wellness return is inferred here [56].

The validity of satisfaction evidence is also influenced by reacting pressure. An employee may become concerned about the chances of their negative answers on a questionnaire ending in their bosses' hands, which is particularly true if questionnaires concern small groups of people, include many demographics and involve medical information. The issue can be solved by ensuring anonymous data collection, restricting the access to a database and allowing the questionnaire to be carried out by third parties. It is crucial to analyze response rates in terms of shifts and employees' categories, as well as considering the relations of missing data to various factors such as overtime, sickness, or dissatisfaction with the program. It is important to conduct an analysis by using probabilistic assumption of missing outcome. Qualitative accounts may shed light on whether the satisfaction responds to the treatment experience, organizational trust or current job situation. These recommendations are particularly relevant for psychosocial risk management and workplace interventions research, where self-report heavily influences outcome measures [57].

The connection between fatigue, error and output requires special care in hazardous production. A strategy that promotes the alertness of employees can reduce errors or near misses. However, measuring only units produced misses a useful safety route. Similarly, shorter cycle times from intensification may increase measured throughput at the cost of fatigue, errors, or morale. The protocol of the balanced measurement should also measure conformance, rework and safety performance, besides productivity. This shall be done in such a way that the interpretation will not conflict with each other. It is important the event rates are calculated using relevant exposure denominators. For near-miss reports, a consideration of the culture of reporting is necessary. Improvements in psychological safety may increase reporting of hazards and near-misses before they increase the rate of incidents. The standards and guidelines that focus on prevention present the conceptual basis for examining this possibility, but the plant study must establish the actual direction and magnitude [58].

A further uncertainty concerns the timing of employment outcomes. It brings further uncertainty on timing of employment outcomes. Changing supervisor communication can affect job satisfaction within weeks, but labour turnover may require the longer observation period. Increased work-ability and changes in health can also be non-linear where recovery, ergonomic learning or organizational adaptation takes time. A trial should not classify an unobserved short-term reduction in turnover as a failure of an intervention, which is primarily designed to reduce strain. Rather, it should identify the relevant endpoint and sufficient follow-up. It is important to report changes from baseline separately from differences between treatment and control. An equally sizeable secular improvement in the comparator group is compatible with a sizeable within-group change and does not, therefore, imply a treatment effect. Randomized studies are essential benchmarks because they reveal the shortcomings of uncontrolled evaluations conducted before-and-after a study [7].

Evidence grades should distinguish demonstrated effectiveness, reasonable plausibility and evidence absence to support implementation decisions. The effects shown need credible comparisons and validated endpoints; plausible interventions may be justified by hazard assessment and occupational safety requirements while effectiveness data is accumulated; absence of evidence is not the same as proven not valuable. The financial cost-benefit ratio need not be positive for preventive risk control under this scheme. It may be possible to pilot other optional programs as expensive expansion is evaluated carefully. By clarifying the distinction between these two ideas, it averts an apparent contradiction between the two. According to evidence from standards, randomized trials and systematic reviews, the most effective interventions are those that are contextually tailored and measured accountably not offering benets in general [59].

DISCUSSION

The key synthesis is that the effectiveness of wellness programs depends on the method of intervention and choosing a specific outcome. In particular, the most rigorous randomised general workplace studies challenge the notion that simply being part of a program from the start will result in either cost savings or better employment results. In contrast, newer studies focused on activities show growth in some health and employability indicators. Out of the 33 surveys conducted, five were successful. Such varying success rates highlight the difference between plausibility and real productivity improvement. These findings are explained partly by population mix, intervention dosage, risk of bias, follow-up, endpoint instruments which denominators are counted and so forth. A plant manager who compares a walking challenge to workstation redesign should avoid ranking interventions based on unrelated self-report survey percentages [7]. The decision should be made based on documented risk exposure, employee priorities and a testable operational hypothesis. Interventions to enhance physical capacities can allow an increase in work tolerance where a musculoskeletal load is limiting. Work redesign targeting fatigue and/or predictability can also assist this all to allow greater productivity or speed of operation. There will not be an automatic improvement in performance of an equipment-limited process unless there is a change in that process [9] .

The key mechanism in job satisfaction is social meaning and usability of the decisions of the organization and not an algorithm. A free resource offered outside paid working hours may be perceived as a benefit by some employees, while night-shift employees may see it as exclusion. The same program may be cast as supportive – when there is worker representation, fair scheduling and confidential support – or surveillance – when it is tied to individual monitoring and performance discipline. The WHO, ISO 45003 and NIOSH make the organizational locus explicit. Synthesis of intervention studies on mental health distinguishes universal prevention from support directed at risk and symptomatic groups to inform access and evaluation. It is important that manufacturing-specific evaluations capture organizational climate, supervisor behaviour and satisfaction as different mechanisms rather than a vague wellness score. To prioritize which types of achievement you want to focus on, you need to consider which ones. Above all, you need to look for the progress in the family activities [12].

An effective future plan must link validated symptom and satisfaction instruments with validated data on participation in clinical trials, occupational exposure measurement and direct production information at the weekly or shift level. A multilevel statistical model can be used to estimate variation between departments and between facilities. A causal interpretation requires appropriate assignment or a defensible quasi-experimental control. Analytical safeguards are feasible. They include a randomized plant-level rollout, preintervention trend testing, blinded outcome extraction where feasible and explicit accounting for maintenance, output mix and absenteeism [60]. When we use predictive machine learning to identify at-risk processes, the out-of-sample accuracy of the predictions does not guarantee that the benefit of intervention will be realized. One needs to test the calibration of the model, the errors on the subgroup of people who respond to the alerts issued and the incremental value if the patient has acted upon alerts issued. Productivity should be measured as conformed output per paid labour hour and accompanied by quality and safety indicators to avoid rewarding risk-taking or speed-up. When it comes to individual illness data, clinical access and production management access should not be combined. In the end, a decision maker needs both an employee-centered effect estimate and a true economic uncertainty interval; the literature reviewed here does not currently offer justification for a universal best programme or a single pooled manufacturing productivity improvement percentage [61].

It is practical in terms of evaluation to get a participatory hazard baseline, select an intervention with a plausible mechanism, confirm access for all shifts and prespecify worker-centered and factory-centered outcomes. An effective operational dataset must identify what is described as a "good unit," using paid labour-hour, rework or scrap, unplanned absence, overtime, staffing, line constraints and exposure changes. To obtain job satisfaction, proper measures need to be taken by using a validated, consistently applied instrument which must be evaluated separately from uptake or engagement with wellness programs. If it is operationally feasible, randomization by line or team is preferred. If this is not possible, matched comparison lines and interrupted time-series approaches should be used. They should record any parallel equipment upgrades, demand changes or process improvements. This is a recommendation for implementation deduced from methodological limitations in the trial and review references and not a proven universal protocol [43].

Limitations of the Study

This critical synthesis review is focused and traceable, not a broad systematic review which is prospectively registered. It focuses on English-language and accessible sources and relevant work from 2020 to October 2026 with foundational trials from 2019. Papers that are indexed in subscription-only databases or journals which cannot be independently verified may therefore be left out from the source set. As the complete counts of database retrieval, duplicate removal and exclusion were not recorded, they are not presented as results. Evidence also varies in terms of population, programme content, follow-up, study design, analytical adjustment and definitions of productivity; notably, work ability, absenteeism and presenteeism are flawed substitutes for objectively measured production output. When synthesis of overlapping trials is done, this creates a false impression of independent replication. The visible advantage might also be blown due to selective outcome reporting and publication bias. The comparative tables retain the numerical evidence at the proper analytical level of each source, with unreported numbers being flagged as such As a result, conclusions are less helpful for estimating a specific numerical return at any particular plant than they are for choosing and assessing interventions [62].

- **Database Limitation:** Subscription-level and exhaustive cross-database retrieval was not completed
- **Methodological Heterogeneity:** Interventions, comparison groups and follow-up periods differ
- **Data and Metric Inconsistency:** Health, satisfaction and production indicators are not interchangeable
- **Limited Generalizability:** Evidence from mixed sectors may not transfer to every factory

CONCLUSION

This review looks at the role of employee wellness programmes in manufacturing productivity and job satisfaction by analyzing experimental workplace research, systematic reviews and occupational-health guidelines. The range of interventions includes traditional behaviour-change programs and physical activity, psychosocial risk prevention, health and safety systems and digital sharing systems. The evidence looked at does not lead to the conclusion that training has a positive impact. It shows that some programmes result in improvement of health behaviour, fitness or work ability. However, this does not mean that this improvement will lead to outputs that are higher, medical spending that is lower, absenteeism that is lower or job satisfaction that is higher in all settings. Several landmark randomized studies show that selection matters and so does causal identification. Recent evidence suggests that some physical and multicomponent interventions could work but that different self-report indicators perhaps should not be used as universal productivity indicators.

In manufacturing facilities, the most effective combination is risk elimination, ergonomic and psychosocial controls, worker involvement, shift-accessible services and confidential individual assistance. All these measures must be motivated by conditions at the workplace as well as the wellbeing of the workers and not by a largely unverified claim of percentage productivity gain. Evaluations of programmes should measure direct labour productivity, quality, safety, work ability, absenteeism, job satisfaction, participation equity and full implementation cost separately. Improvements should be indicated with reference to a defined comparison group and uncertainty and with controls for production mix, staffing, maintenance and business-cycle changes? Though machine-learning or wearable technologies that speed up delivery of services or increase inclusiveness may assist in prediction, it is important not to confuse prediction performance with evidence that any subsequent intervention is effective.

The main focus should be on multicentre manufacturing trials that are well powered and credible phased-rollout studies that measure baseline work conditions, intervention fidelity, worker privacy, longer-term maintenance and prespecified economic evaluation. To avoid double counting of reviews and constituent studies, every effect size should retain its source specific definition in transparent reporting. Through the delineation of plausible causal mechanisms from benefits demonstrated independently, a robust framework is provided within the present review for selecting wellness components appropriate for plants as well as evaluating whether the assessed wellness components truly enhance employee experiences as well as operational outcomes.

REFERENCES

1. Afin, S. "The role of ergonomics in lean manufacturing: enhancing productivity, worker well-being and operational performance." 2026.
2. Malik, A. *et al*. "Mental health at work: WHO guidelines." *World Psychiatry*, vol. 22, no. 2, 2023, pp. 331–332. https://doi.org/10.1002/wps.21094.
3. Tamers, S.L. *et al*. "Total worker health® 2014–2018: The novel approach to worker safety, health and well-being evolves." *International Journal of Environmental Research and Public Health*, vol. 16, no. 3, 2019, pp. 321. https://doi.org/10.3390/ijerph16030321.
4. Noben, C.Y. *et al*. "Quality appraisal of generic self-reported instruments measuring health-related productivity changes: A systematic review." *BMC Public Health*, vol. 14, no. 1, 2014, pp. 115. https://doi.org/10.1186/1471-2458-14-115.
5. Sakuraya, A. *et al*. "Effects of participatory organizational interventions on mental health and work performance: A Systematic review and meta-analysis." *Journal of Occupational Health*, vol. 68, no. 1, 2026, uiag024. https://doi.org/10.1093/joccuh/uiag024.
6. Kekkonen, P. *et al*. "Effectiveness of work ability interventions on productivity: A systematic review." *Journal of Occupational Health*, vol. 67, no. 1, 2025, uiaf051. https://doi.org/10.1093/joccuh/uiaf051.
7. Song, Z. and K. Baicker. "Effect of a workplace wellness program on employee health and economic outcomes: A randomized clinical trial." *JAMA*, vol. 321, no. 15, 2019, pp. 1491. https://doi.org/10.1001/jama.2019.3307.
8. Jones, D. *et al*. "What do workplace wellness programs do? Evidence from the illinois workplace wellness study." *The Quarterly Journal of Economics*, vol. 134, no. 4, 2019, pp. 1747–1791. https://doi.org/10.1093/qje/qjz023.
9. Marin-Farrona, M. *et al*. "Effectiveness of worksite wellness programs based on physical activity to improve workers' health and productivity: A systematic review." *Systematic Reviews*, vol. 12, no. 1, 2023, pp. 87. https://doi.org/10.1186/s13643-023-02258-6.
10. Støren, P.G. and H. Grønningsæter. "Do worksite health promotion programs (whpp) influence presenteeism among employees? A systematic review." *Work*, vol. 77, no. 1, 2024, pp. 85–102. https://doi.org/10.3233/WOR-220115.

11. Javanmardi, S. *et al*. "Effectiveness of workplace health promotion programs for industrial workers: A systematic review." *BMC Public Health*, vol. 25, no. 1, 2025, pp. 168. https://doi.org/10.1186/s12889-025-21365-8.
12. Miguel, C. *et al*. "Universal, selective and indicated interventions for supporting mental health at the workplace: An umbrella review of meta-analyses." *Occupational and Environmental Medicine*, vol. 80, no. 4, 2023, pp. 225–236. https://doi.org/10.1136/oemed-2022-108698.
13. Page, M.J. *et al*. "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews." *BMJ*, 2021, n71. https://doi.org/10.1136/bmj.n71.
14. Kwasnitschka, D. *et al*. "Effects of feedback in manufacturing: A field experiment using smartwatch technology." *Journal of Operations Management*, vol. 70, no. 6, 2024, pp. 933–956. https://doi.org/10.1002/joom.1305.
15. Ng Corrales, L.D.C. *et al*. "Overall equipment effectiveness: Systematic literature review and overview of different approaches." *Applied Sciences*, vol. 10, no. 18, 2020, pp. 6469. https://doi.org/10.3390/app10186469.
16. Spector, P.E. "Measurement of human service staff satisfaction: Development of the job satisfaction survey." *American Journal of Community Psychology*, vol. 13, no. 6, 1985, pp. 693–713. https://doi.org/10.1007/BF00929796.
17. World Health Organization. *WHO Guidelines on Mental Health at Work*. 2022. https://www.who.int/publications/b/63912.
18. Jarden, R.J. *et al*. "Wellbeing measures for workers: A systematic review and methodological quality appraisal." *Frontiers in Public Health*, vol. 11, 2023, pp. 1053179. https://doi.org/10.3389/fpubh.2023.1053179.
19. Moher, D. *et al*. "CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomized trials." *BMJ*, vol. 340, 2010, c869. https://doi.org/10.1136/bmj.c869.
20. Amrhein, V. *et al*. "Scientists rise up against statistical significance." *Nature*, vol. 567, no. 7748, 2019, pp. 305–307. https://doi.org/10.1038/d41586-019-00857-9
21. Rostami, M. *et al*. "Assessing the effectiveness of an ergonomics intervention program with a participatory approach: Ergonomics settlement in an Iranian steel industry." *International Archives of Occupational and Environmental Health*, vol. 95, no. 5, 2022, pp. 953–964. https://doi.org/10.1007/s00420-021-01811-x.
22. Von Thiele Schwarz, U. *et al*. "How to design, implement and evaluate organizational interventions for maximum impact: The Sigtuna principles." *European Journal of Work and Organizational Psychology*, vol. 30, no. 3, 2021, pp. 415–427. https://doi.org/10.1080/1359432X.2020.1803960.
23. Tamminga, S.J. *et al*. "Individual-level interventions for reducing occupational stress in healthcare workers." *Cochrane Database of Systematic Reviews*, vol. 2023, no. 5, 2023. https://doi.org/10.1002/14651858.CD002892.pub6.
24. Amoadu, M. *et al*. "Influence of psychosocial safety climate on occupational health and safety: A scoping review." *BMC Public Health*, vol. 23, no. 1, 2023, pp. 1344. https://doi.org/10.1186/s12889-023-16246-x.
25. Anger, W.K. *et al*. "Effectiveness of total worker health interventions." *Journal of Occupational Health Psychology*, vol. 20, no. 2, 2015, pp. 226–247. https://doi.org/10.1037/a0038340.
26. Blattman, C. and L. Ralston. "Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs." *SSRN*, 2015. https://doi.org/10.2139/ssrn.2622220.
27. Both, C.M. and M. Wilhelm. "Supply chain sustainability risk management in the era of mandatory due diligence: A literature review." *Journal of Purchasing and Supply Management*, vol. 32, no. 2, 2026, 101083. https://doi.org/10.1016/j.pursup.2025.101083.
28. Rugulies, R. *et al*. "Work-related causes of mental health conditions and interventions for their improvement in workplaces." *The Lancet*, vol. 402, no. 10410, 2023, pp. 1368–1381. https://doi.org/10.1016/S0140-6736(23)00869-3.
29. Moe-Byrne, T. *et al*. "Effectiveness of tailored digital health interventions for mental health at the workplace: A systematic review of randomized controlled trials." *PLOS Digital Health*, vol. 1, no. 10, 2022, e0000123. https://doi.org/10.1371/journal.pdig.0000123.
30. Vickers, A.J. *et al*. "Net benefit approaches to the evaluation of prediction models, molecular markers and diagnostic tests." *BMJ*, 2016, i6. https://doi.org/10.1136/bmj.i6.
31. Collins, G.S. *et al*. "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods." *BMJ*, vol. 385, 2024, e078378. https://doi.org/10.1136/bmj-2023-078378.
32. Pehlevan, E. and Ö. Şevgin. "Effect of exercise given to factory workers with ergonomics training on pain and functionality: A randomized controlled trial." *WORK: A Journal of Prevention, Assessment & Rehabilitation*, vol. 78, no. 1, 2024, pp. 195–205. https://doi.org/10.3233/WOR-230663.

33. Skamagki, G. *et al*. "Workplace interventions for chronic musculoskeletal disorders: A systematic review." *BMJ Open*, vol. 16, no. 6, 2026, e115276. https://doi.org/10.1136/bmjopen-2025-115276.
34. Fritz Lerche, A. *et al*. "Effectiveness of a Goldilocks Work Intervention to promote musculoskeletal health among industrial workers – A cluster randomized controlled trial." *Journal of Safety Research*, vol. 90, 2024, pp. 392–401. https://doi.org/10.1016/j.jsr.2024.05.006.
35. Jin, X. *et al*. "Ergonomic interventions to improve musculoskeletal disorders among vehicle assembly workers: A one-year longitudinal study." *BMC Public Health*, vol. 25, no. 1, 2025, pp. 824. https://doi.org/10.1186/s12889-025-21798-1.
36. Song, Z. and K. Baicker. "Health and economic outcomes up to three years after a workplace wellness program: A randomized controlled trial." *Health Affairs*, vol. 40, no. 6, 2021, pp. 951–960. https://doi.org/10.1377/hlthaff.2020.01808.
37. Smith, N.D. *et al*. "Digital performance management: an evaluation of manufacturing performance management and measurement strategies in an industry 4.0 context." *Machines*, vol. 12, no. 8, 2024, pp. 555. https://doi.org/10.3390/machines12080555.
38. Burr, H. *et al*. "The third version of the copenhagen psychosocial questionnaire." *Safety and Health at Work*, vol. 10, no. 4, 2019, pp. 482–503. https://doi.org/10.1016/j.shaw.2019.10.002.
39. Lutz, N. *et al*. "Cost-effectiveness and cost-benefit of worksite health promotion programs in Europe: A systematic review." *European Journal of Public Health*, vol. 29, no. 3, 2019, pp. 540–546. https://doi.org/10.1093/eurpub/cky269.
40. Wasserstein, R.L. and N.A. Lazar. "The ASA statement on p-values: Context, process and purpose." *The American Statistician*, vol. 70, no. 2, 2016, pp. 129–133. https://doi.org/10.1080/00031305.2016.1154108.
41. Rojatz, D. *et al*. "Factors influencing workplace health promotion intervention: A qualitative systematic review." *Health Promotion International*, daw015, 2016. https://doi.org/10.1093/heapro/daw015.
42. Egan, M. *et al*. "The psychosocial and health effects of workplace reorganisation. 1. A systematic review of organisational-level interventions that aim to increase employee control." *Journal of Epidemiology and Community Health*, vol. 61, no. 11, 2007, pp. 945–954. https://doi.org/10.1136/jech.2006.054965.
43. Skivington, K. *et al*. "A new framework for developing and evaluating complex interventions: Update of medical research council guidance." *BMJ*, n2061, 2021. https://doi.org/10.1136/bmj.n2061.
44. National Institute for Occupational Safety and Health. "About the total worker health approach." *Centers for Disease Control and Prevention*, 2024. https://www.cdc.gov/niosh/twh/about/.
45. International Organization for Standardization. *Occupational Health and Safety Management—Psychological Health and Safety at Work—Guidelines for Managing Psychosocial Risks*. ISO 45003:2021, 2021. https://www.iso.org/standard/64283.html.
46. Welch, V.A. *et al*. "CONSORT-equity 2017 extension and elaboration for better reporting of health equity in randomized trials." *BMJ*, j5085, 2017. https://doi.org/10.1136/bmj.j5085.
47. Aust, B. *et al*. "How effective are organizational-level interventions in improving the psychosocial work environment, health and retention of workers? A systematic overview of systematic reviews." *Scandinavian Journal of Work, Environment & Health*, vol. 49, no. 5, 2023, pp. 315–329. https://doi.org/10.5271/sjweh.4097.
48. Stratton, E. *et al*. "Digital m health interventions for employees: Systematic review and meta-analysis of their effects on workplace outcomes." *Journal of Occupational & Environmental Medicine*, vol. 63, no. 8, 2021, pp. e512–e525. https://doi.org/10.1097/JOM.0000000000002267.
49. Mendes, D. *et al*. "Wearable technologies in occupational safety and health: A systematic review and a human-centered implementation model." *Applied Sciences*, vol. 16, no. 10, 2026, pp. 4715. https://doi.org/10.3390/app16104715.
50. Saad Mahdi *et al*. "The role of employee wellness programs in sustainable workplace productivity." *Environment and Social Psychology*, vol. 10, no. 12, 2025. https://doi.org/10.59429/esp.v10i12.3993.
51. Daniels, K. *et al*. "Implementing practices focused on workplace health and psychological wellbeing: a systematic review." *Social Science & Medicine*, vol. 277, 2021, pp. 113888. https://doi.org/10.1016/j.socscimed.2021.113888.
52. Moore, G.F. *et al*. "Process evaluation of complex interventions: Medical research council guidance." *BMJ*, vol. 350, 2015, pp. h1258. https://doi.org/10.1136/bmj.h1258.
53. NIOSH. *Fundamentals of Total Worker Health® Approaches: Essential Elements for Advancing Worker Safety, Health and Well-Being*. 2017. https://www.cdc.gov/niosh/docs/2017-112/.
54. Wiltsey Stirman, S. *et al*. "The FRAME: An expanded framework for reporting adaptations and modifications to evidence-based interventions." *Implementation Science*, vol. 14, no. 1, 2019, pp. 58. https://doi.org/10.1186/s13012-019-0898-y.

55. Puri, P. "Mixed methods design for the evaluation of development projects across baseline, midline and endline." *Research and Innovation*, vol. 2, no. 2, 2026, pp. 129–143. https://doi.org/10.11648/j.ri.20260202.14.
56. Lensberg, B.R. *et al*. "Challenges in measuring and valuing productivity costs and their relevance in mood disorders." *Clinico Economics and Outcomes Research*, 2013, pp. 565. https://doi.org/10.2147/CEOR.S44866.
57. Strudwick, J. *et al*. "Workplace mental health screening: A systematic review and meta-analysis." *Occupational and Environmental Medicine*, vol. 80, no. 8, 2023, pp. 469–484. https://doi.org/10.1136/oemed-2022-108608.
58. Techera, U. *et al*. "Causes and consequences of occupational fatigue: Meta-analysis and systems model." *Journal of Occupational & Environmental Medicine*, vol. 58, no. 10, 2016, pp. 961–973. https://doi.org/10.1097/JOM.0000000000000837.
59. Gottfredson, D.C. *et al*. "Standards of evidence for efficacy, effectiveness and scale-up research in prevention science: Next generation." *Prevention Science*, vol. 16, no. 7, 2015, pp. 893–926. https://doi.org/10.1007/s11121-015-0555-x.
60. Hemming, K. *et al*. "The stepped wedge cluster randomized trial: Rationale, design, analysis and reporting." *BMJ*, vol. 350, 2015, pp. h391. https://doi.org/10.1136/bmj.h391.
61. Riley, R.D. *et al*. "Evaluation of clinical prediction models (Part 2): How to undertake an external validation study." *BMJ*, vol. 384, 2024, e074820. https://doi.org/10.1136/bmj-2023-074820.
62. Snyder, H. "Literature review as a research methodology: An overview and guidelines." *Journal of Business Research*, vol. 104, 2019, pp. 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039.