

A joint modeling framework for time-to-fatigue prediction with a single wearable sensor biomarker

Received: 18 March 2025

Accepted: 18 February 2026

Published online: 06 March 2026

Cite this article as: Lu L., Sedighi-Maman Z. & Cavuoto L. A joint modeling framework for time-to-fatigue prediction with a single wearable sensor biomarker. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-41249-0>

Lin Lu, Zahra Sedighi-Maman & Lora Cavuoto

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A Joint Modeling Framework for Time-to-Fatigue Prediction with a Single Wearable Sensor Biomarker

Lin Lu¹, Zahra Sedighi-Maman^{2,*}, and Lora Cavuoto³

¹Charles F. Dolan School of Business, Fairfield University, Connecticut, CT 06824, USA

^{2,*}Corresponding Author, Robert B. Willumstad School of Business, Adelphi University, Garden City, NY 11530, USA, Email: zmaman@adelphi.edu

³Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260, USA

ABSTRACT

Physical fatigue is a prevalent issue in the U.S. workforce and has resulted in reduced productivity and a higher incidence of workplace accidents. Being able to estimate time to fatigue facilitates proactive interventions before fatigue onset. Existing research has primarily focused on the fatigue state, intensity, and wearable sensor characteristics while there remains a gap in predicting time to fatigue and identifying indicative biomarkers that can dynamically signal impending fatigue risks. In this study, we utilize simulated manufacturing work tasks and employ a joint modeling framework that links a longitudinal biomarker to predict individual time-to-fatigue outcomes. Across two case studies with differing task demands and durations, the joint model with a single sensor trajectory consistently outperformed baseline methods, including Kaplan-Meier estimates and Cox proportional hazards models, demonstrating its effectiveness for dynamic fatigue prediction in simulated manufacturing tasks while accounting for variations across tasks and individual sensor responses. Although predictive features may vary by task, our results suggest that, within the contexts examined, jerk-based features—identified in previous studies, including our own, as indicators of fatigue states—may also serve as predictors of time to fatigue. These findings advance the understanding of fatigue progression and offer foundational insights for supporting timely interventions in future applied settings. This work highlights the potential of online wearable monitoring and interpretable statistical methods to inform proactive occupational safety strategies and worker well-being initiatives.

Introduction

Physical fatigue refers to a decline in strength, physical capacity, or performance as result of work activities¹. It is a prevalent issue in physically demanding industrial environments^{2,3}, contributing to higher rates of injuries, errors, deficits, and process disruptions⁴⁻⁶. Under the Industry 4.0 paradigm, increased automation, rising efficiency standards, and greater human-machine interactions have intensified work demands, often leading to faster work pace and exacerbating physical fatigue^{7,8}.

Estimating time to fatigue (TTF) and introducing timely interventions can be essential for conquering these challenges. From a managerial perspective, being able to predict when workers are likely to experience fatigue enables the implementation of proactive strategies, such as adjusted work-rest schedules and job rotations, to reduce injury risk and sustain productivity. From a modeling standpoint, relying solely on self-reported surveys or direct observations to measure TTF is prone to subjectivity and can be labor-intensive⁹. Notably, advancements in wearable sensors overcome many of these limitations by providing objective and continuous data that can be leveraged for scalable and real-time fatigue monitoring.

Over the past decade, physical fatigue monitoring studies using wearable sensors have primarily focused on three outcomes: (A) fatigue states, where subjects' categorized fatigue states are classified at discrete time points¹⁰⁻¹⁴; (B) fatigue levels, which estimates continuous fatigue measurements, such as rate of perceived exertion (RPE), over time^{3,15,16}; and (C) fatigue-associated sensor variables, which describe the characteristics of sensor features (also referred to as biomarkers in this context) by fatigue states¹⁷⁻¹⁹. Although these studies have advanced fatigue assessment, they do not explicitly estimate TTF, limiting their applicability for proactive fatigue management. Specifically, fatigue classification models are usually cross-sectional, constraining their ability to utilize temporal variations from wearable sensors¹⁷. Fatigue intensity prediction models, relying on self-reported measures such as RPE, are inherently subjective and may introduce variability due to individual differences in perception, where small variations may not always reflect meaningful physiological differences¹⁴. Studies on fatigue-associated sensor variables have largely been exploratory rather than providing real-time, actionable monitoring, which constrains their ability to provide immediate feedback to workers during tasks¹⁸.

Beyond wearable sensor-based research, relevant studies on TTF estimation have examined endurance time using explanatory mathematical models²⁰⁻²³. These models typically incorporate time-fixed task- or subject-related factors, such as force levels, postures, or demographic characteristics. While such models are useful for optimizing work designs, their applicability is often

limited to highly controlled lab environments and specific task settings and does not account for task dynamics or individual differences, reducing their suitability for real-time and individualized predictions.

These research gaps underscore the need for TTF prediction models that can incorporate biomarkers to allow consideration of task dynamics and real-time updates of human conditions⁸. In response, this study examines joint modeling (JM) of time-to-event and longitudinal outcomes as a promising framework. JM techniques are suitable and have been applied in settings where subjects are followed over time with repeated measurements of biomarkers for monitoring the progression of a disease or medical condition²⁴. Unlike time-dependent Cox models of which assumptions include exogeneity and no measurement error, JM is particularly useful when longitudinal covariates are endogenous or exhibit complex dependencies with the survival outcome²⁵. Existing applications of JM include tracking cognitive decline in Alzheimer's disease²⁶, evaluating survival and treatment effects in chronic liver disease²⁴, assessing mortality risk in severe aortic stenosis based on brain natriuretic peptide levels²⁷, and modeling heart failure-related mortality²⁸. In comparison, while sharing a similar focus on longitudinal monitoring of a health condition progression through endogenous biomarker measurements, the application of JM to occupational fatigue remains largely unexplored. By explicitly linking the trajectory of a sensor biomarker to the likelihood of fatigue over time, JM shows the potential of dynamic fatigue prediction while accommodating occasional deviations in measurement times common in sensor data and accounting for individual variability²⁹.

To address research gaps in TTF prediction and leverage the strengths of JM, this study examines the feasibility of JM on predicting time to occupational physical fatigue. Specifically, we focus on evaluating the predictive capability of a single sensor feature per task to enhance model interpretability, streamline practical monitoring, and facilitate a clear examination of sensor-specific biomarker relationships with TTF. Through analyses of two independently conducted case studies involving simulated manufacturing work tasks, this study explores the following research questions:

- **How effectively can JM predict TTF utilizing a single sensor feature?** We apply both classical survival analysis and JM approaches to predict subjects' TTF during the task and evaluate predictive performance using the concordance index (CND) and mean absolute percentage error (MAPE) to assess model discrimination and accuracy.
- **Which sensor features are most predictive of TTF within a given work task, and how do these sensor features associate with fatigue survival probabilities over time?** We examine sensor features from inertial measurement unit (IMU) sensors and heart rate monitors that capture fatigue-related biomechanical and physiological changes during task sessions. The most predictive sensor features of TTF are identified based on predictive performance metrics and statistical significance. The associations between sensor features and fatigue survival probabilities are quantified using hazard ratios and further illustrated through dynamic predictions and comparative analyses.

By addressing these questions, this proof-of-concept study offers insights into the potential role of sensor-based monitoring for understanding fatigue progression and informing proactive intervention strategies aimed at improving workplace safety and efficiency. Additionally, it contributes to the broader literature on time-to-event modeling with wearable technology, demonstrating the potential of this approach for real-time monitoring and decision-making.

Results

We conducted TTF predictions for subjects performing simulated manufacturing work tasks in two independent case studies. Across both case studies, TTF was defined as the elapsed time from the start of the task to the first occurrence at which a predefined task-specific fatigue threshold was reached. Case Study 1 includes manual material handling (MMH) and supply insertion (SI) tasks, both of which captured the same set of sensor features. Case Study 2 involves bottle picking (BP) under four experimental conditions, varying by bottle mass (kg) and picking rate, measured in bottles per minute (bott/min). In these two case studies, we evaluated predictive performance using CND and MAPE based on leave-two-subject-out cross-validation. In addition, we illustrated dynamic predictions by visualizing concurrent changes in the best performing sensor features and survival probabilities during task execution.

Case Study 1: MMH and SI Tasks

Predictive Performance

During the 180-minute simulated MMH and SI sessions, valid data were obtained from 15 of 19 subjects for the MMH task and 12 for the SI task. The remaining sessions were excluded due to protocol deviations or sensor malfunctions. Among the 27 valid task sessions, subjects reached a fatigued state in 17, defined as a Borg RPE ≥ 15 . Consequently, 37% of the sessions (10 out of 27) were censored, as subjects did not reach the predefined fatigue threshold within the session duration. In the MMH and SI tasks, TTF was calculated as the time elapsed from task initiation to the first instance in which a subject rated their fatigue as 15 or higher on the Borg 6–20 scale.

Of all the sensor features, five—*wrist.jerk.sd*, *wrist.jerk.mean*, *wrist.acc.sd*, *ankle.zposture.sd*, and *hip.acc.sd*—showed an absolute point-biserial correlation of $|r| \geq 0.2$ with time-varying fatigue states in more than half of the 171 leave-two-subject-out cross-validation (i.e., $\binom{19}{2}$) folds, and were selected as candidate time-varying covariates for JM. Each of these features also demonstrated statistical significance ($p < 0.05$) in univariate time-dependent Cox models across more than half of the folds, supporting the association with TTF. When each candidate was examined individually in JM, *wrist.jerk.mean* achieved the highest average CND and the lowest MAPE across folds, identifying it as the most predictive feature. The best performing model followed a random intercepts and random slopes (RIRS) form in the linear mixed-effects submodel, incorporating both the *wrist.jerk.mean* value and its slope in JM, without requiring the inclusion of the *Task* factor (i.e., MMH or SI). Across the 171 cross-validation training folds, both the value and slope of *wrist.jerk.mean* from the best performing JM were significant ($p \leq 0.05$) in a majority of folds, with the value significant in 167 folds and the slope in 138 folds—demonstrating greater consistency for the value component.

Table 1 compares the prediction performance of the baseline and JM approaches towards TTF. Among the baseline models, Kaplan-Meier (KM) with the *Task* factor (time-fixed covariate) achieved the highest test-fold CND as 56.24% and a MAPE of 29.56%, and its training-fold metrics were 54.45% and 21.87%, respectively. In contrast, incorporating *wrist.jerk.mean* as a longitudinal covariate in JM with six or seven measurements (i.e., at follow-up times of 60 and 70 minutes, based on 10-minute intervals aligned with RPE measurements), while excluding the *Task* factor, increased both test and train CND to above 72% and reduced MAPE to below 19%.

Table 1. Case Study 1: Predictive performance of baseline and JM approaches

Models		KM	KM	Cox	JM: <i>wrist.jerk.mean</i>		JM: <i>wrist.jerk.mean</i>	
Time-fixed covariate		1	Task		1		Task	
After the n th measurement interval		-			6th	7th	6th	7th
Train	CND%	51.98	54.45	54.04	77.48	76.60	76.63	76.13
	MAPE%	23.58	21.87	22.04	18.00	18.50	18.47	18.65
Test	CND%	53.55	56.24	51.26	72.71	73.36	71.34	72.31
	MAPE%	29.94	29.56	26.88	18.57	18.38	19.64	19.33

Dynamic Predictions with a Single Time-varying Sensor Feature

The hazard ratio for *wrist.jerk.mean* is 1.126 ($p = 0.017$, significant at $p < 0.05$), based on the simplified best-performing JM applied to the complete dataset. The simplified model still adopts RIRS in the longitudinal process, but excludes the slope of *wrist.jerk.mean* in its JM parameterization to isolate the effect of the biomarker value. This specification considers the greater predictive relevance in the sensor feature's value component and allows a clearer interpretation of its direct association with fatigue progression. The result indicates that each one-unit increase in *wrist.jerk.mean* corresponds to a 12.6% higher hazard of fatigue onset. To contextualize this effect, *wrist.jerk.mean* values across sessions ranged from 11.57 to 16.41 (25th to 75th percentile), which corresponds to an estimated 78% higher fatigue hazard between these quartiles.

Figure 1 illustrates dynamic predictions of median fatigue survival probabilities and associated 95% confidence intervals (CIs) for Subject 2 during the SI session. The subject's *wrist.jerk.mean* exhibits a clear increase from the sixth to the seventh measurement, accompanied by a steeper decline in fatigue survival probabilities. Moreover, Figure 2 compares predicted fatigue survival probabilities for Subject 2 and 4 during the SI task, evaluated after the same two measurement points, each followed by prediction windows of 40, 70, and 100 minutes. Survival probabilities for Subject 2 decreased more rapidly across these windows than for Subject 4, indicating an earlier predicted fatigue onset. By the end of the task session, Subject 2 ultimately reported fatigue, while Subject 4 did not.

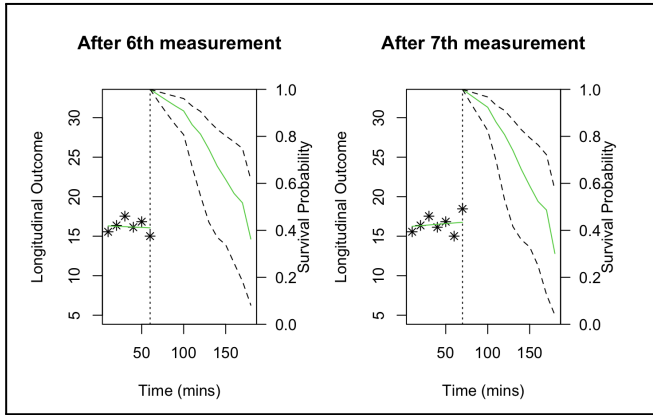


Figure 1. Predicted median fatigue survival probabilities (green curve) for Subject 2 during the SI task, shown with 95% CIs (dashed lines). Predictions were evaluated after the 6th and 7th measurements of *wrist.jerk.mean* (asterisks), corresponding to 60 and 70 minutes of follow-up, respectively.

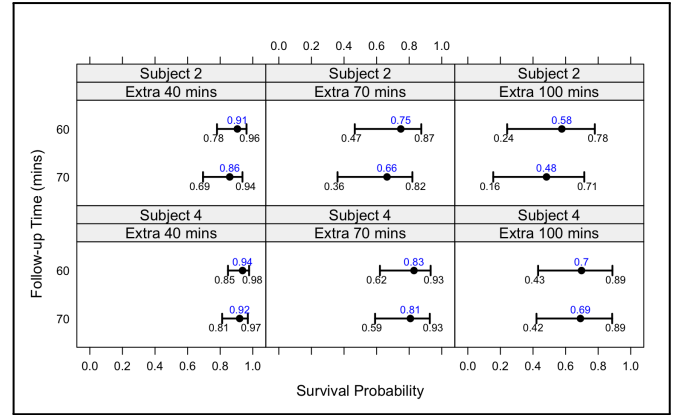


Figure 2. Predicted median fatigue survival probabilities (dots) for Subject 2 and 4 during the SI task, shown with 95% CIs (horizontal bars). Estimates were generated at follow-up times 60 and 70 minutes (rows), with predictions for the next 40, 70, and 100 minutes ahead (columns).

Sensor Trajectory Patterns

To visualize how the most predictive feature evolves over time in Case Study 1, Figure 3 shows the mean (solid lines) and median (dashed lines) of *wrist.jerk.mean* for fatigued (red) and non-fatigued (blue) participants during the MMH and SI tasks. Across both tasks, the fatigued group generally exhibits higher jerk values, reflecting less smooth and less controlled movements compared to the non-fatigued group. This group-level pattern aligns with the joint model results, where elevated *wrist.jerk.mean* is associated with increased fatigue risk. Note that after 120 minutes, only two MMH participants remain at risk, which contributes to the observed drop in the group mean and median around that time.

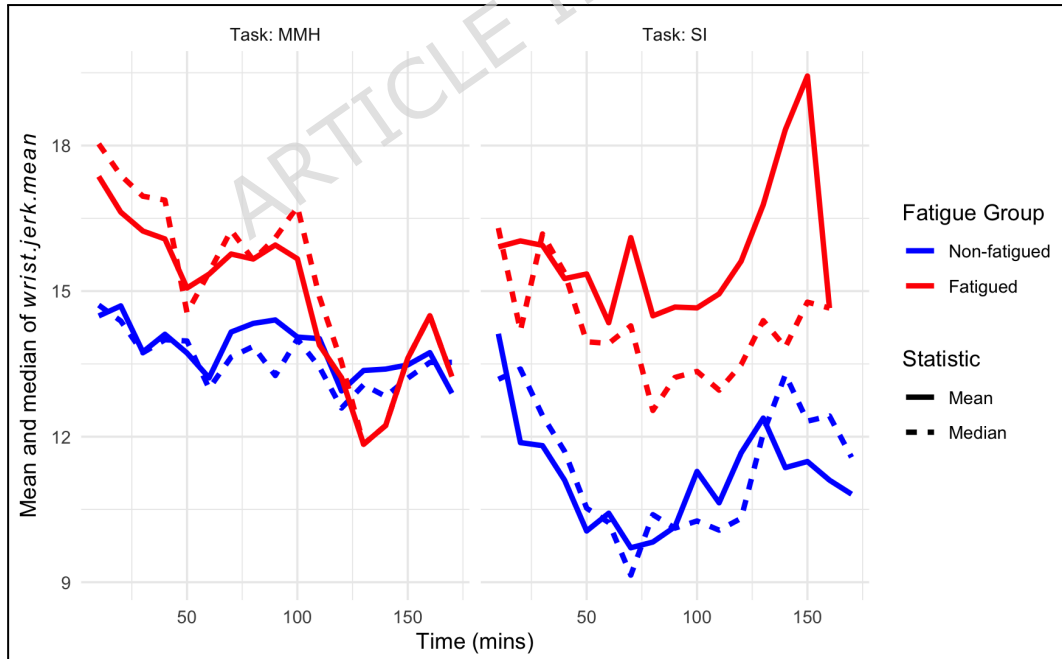


Figure 3. Mean and median values of *wrist.jerk.mean* over time by task and fatigue group, aggregated across subjects

Case Study 2: BP task

Predictive Performance

During the 45-minute experimental sessions, 14 subjects participated in the BP task. Five subjects completed all four sessions, while the remaining subjects completed two or three sessions. Invalid sessions were excluded due to protocol deviations or

sensor malfunctions. In total, 45 task sessions' data were obtained across four experimental conditions, which varied in bottle mass and picking rate: 2.5 kg $\times$ 15/10/5 bott/min and 1.5 kg $\times$ 15 bott/min. Among these, subjects in 30 sessions reached a fatigued state, defined as Borg CR-10 RPE $\geq$ 4, resulting in a censoring rate of 33.3% (15 of 45 sessions). In the BP task, TTF was calculated as the time elapsed from the beginning of the task until a subject rated their fatigue as 4 or higher on the Borg CR-10 scale. Six sensor features—*wrist.acc.mad*, *wrist.acc.rms*, *wrist.acc.sd*, *wrist.jerk.perc25*, *upperarm.acc.perc25*, and *upperarm.jerk.perc25*—exhibited point-biserial correlations of $|r| \geq 0.2$ with time-varying fatigue states in more than 91 leave-two-subject-out cross-validation folds (i.e., $\binom{19}{2}$). These sensor features were also statistically significant ($p < 0.05$) in univariate time-dependent Cox models in more than half of the folds, confirming their associations with TTF. Each of these candidate sensor features was subsequently fitted as a time-varying covariate in separate joint models. Among them, *wrist.jerk.perc25* demonstrated the best performance. The optimal JM configuration uses an RIRS linear mixed-effects submodel, excludes the *Experimental condition* factor in the Cox submodel, and incorporates the biomarker slope alone in parameterization. The slope of *wrist.jerk.perc25* was significant ($p \leq 0.05$) in all but four cross-validation folds (87 out of 91 folds), confirming its predictive relevance.

Table 2 summarizes the TTF predictive performance for the BP task. Among the baseline models, the Cox model with the *Experimental condition* factor (time-fixed covariate) achieved a test-fold CND% of 65.32 and a corresponding MAPE of 69.87%, while in the train fold, it attained CND and MAPE values of 64.25% and 56.6%, respectively. In comparison, the best performing JM using *wrist.jerk.perc25* as a biomarker with six or seven measurements (i.e., at follow-up times of 15 and 17.5 minutes, based on 2.5-minute intervals used for sensor feature calculation) reached a CND% above 72% and a MAPE mostly below 30% in both test and train folds.

Table 2. Case Study 2: Predictive performance of baseline and JM approaches

Models		KM / Cox	KM	Cox	JM: <i>Wrist.jerk.perc25</i>		JM: <i>Wrist.jerk.perc25</i>	
Time-fixed covariate		1	Experimental condition		1		Experimental condition	
After the n th measurement interval					6th	7th	6th	7th
Train	CND%	50.69	61.51	64.25	72.81	75.47	69.61	74.93
	MAPE%	71.32	46.06	56.6	29.09	24.11	32.64	26.15
Test	CND%	50.95	59.67	65.32	72.55	74.90	67.74	72.84
	MAPE%	84.60	61.93	69.87	30.88	22.92	36.13	25.50

Dynamic Predictions with a Single Time-varying Sensor Feature

The hazard ratio for the slope of *wrist.jerk.perc25* is 0.367 ($p = 0.01$, significant at $p < 0.05$) based on the best performing JM fitted to the complete dataset, revealing that an one-unit increase in the scaled slope of *wrist.jerk.perc25* (slope $\times$ 100) is associated with a 63.3% reduction in the hazard of fatigue. To contextualize this effect, the relationship between the slope of *wrist.jerk.perc25* and fatigue hazard was sensitive. A small increase in the slope of 0.007 was associated with a 50% reduction in the hazard of fatigue.

The above hazard ratio value shows that a greater positive slope of *wrist.jerk.perc25* is linked to delayed fatigue onset, as illustrated in Figure 4. During the experimental condition of 2.5 kg $\times$ 5 bott/min, the subject 20 exhibited a clear increase in the slope of *wrist.jerk.perc25* between the sixth and seventh measurements, which appeared to slow the decline in the predicted fatigue survival probability for the remainder of the session. Furthermore, Figure 5 compares predicted survival probability for Subject 20 and 34 under the experimental condition of 2.5 kg $\times$ 10 bott/min, evaluated after the same two measurement intervals, each followed by prediction windows of 10, 15, and 25 minutes. The survival probabilities for Subject 34 declined at a faster rate over time compared to Subject 20. By the end of the task session, Subject 34 ultimately reported fatigue, while Subject 20 did not.

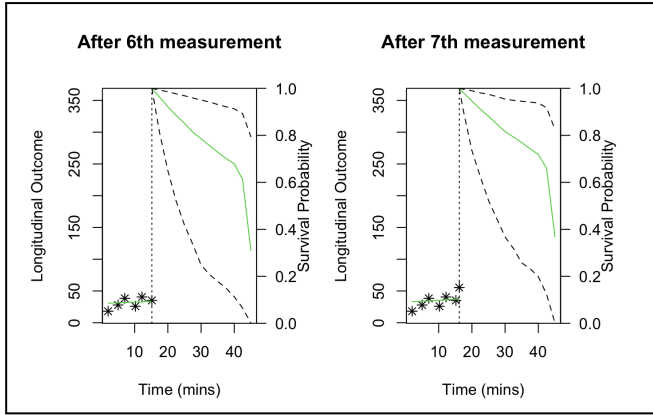


Figure 4. Predicted median fatigue survival probabilities (green curve) for Subject 20 in the 2.5 kg $\times$ 10 bott/min BP task, shown with 95% CIs (dashed lines). Predictions were evaluated after the 6th and 7th measurements of *wrist.jerk.perc25* (asterisks), corresponding to 15 and 17.5 minutes of follow-up, respectively.

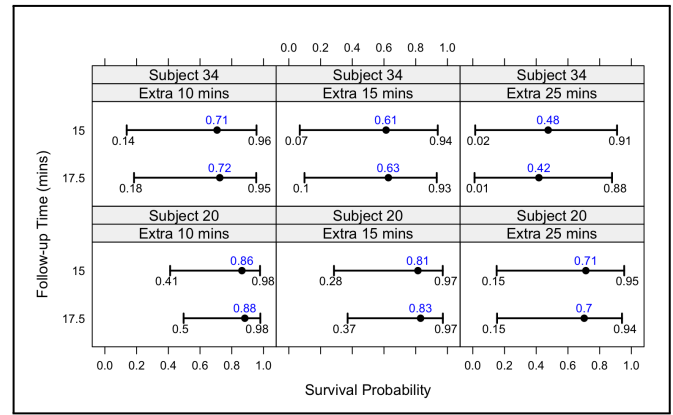


Figure 5. Predicted median fatigue survival probabilities (dots) for Subject 20 and 34 in the 2.5 kg $\times$ 5 bott/min BP task, shown with 95% CIs (horizontal bars). Estimates were generated at follow-up times of 15 and 17.5 minutes (rows), with predictions for the next 10, 15, and 25 minutes ahead (columns).

Sensor Trajectory Patterns

Figure 6 presents the trend lines of *wrist.jerk.perc25* under different experimental conditions of the BP task, highlighting the overall slope differences. In particular, under the 1.5 kg $\times$ 15 bott/min and 2.5 kg $\times$ 15 bott/min conditions, the fatigued groups exhibit a downward trend over time, indicating diminished physical responsiveness, whereas non-fatigued subjects generally show an upward trend with continued task repetitions. These patterns align with the JM results from Case Study 2, which indicate that an increasing *wrist.jerk.perc25* slope is significantly associated with a lower hazard of fatigue, suggesting that subjects maintained sufficient physical energy to compensate for workload demands, thereby delaying fatigue onset.

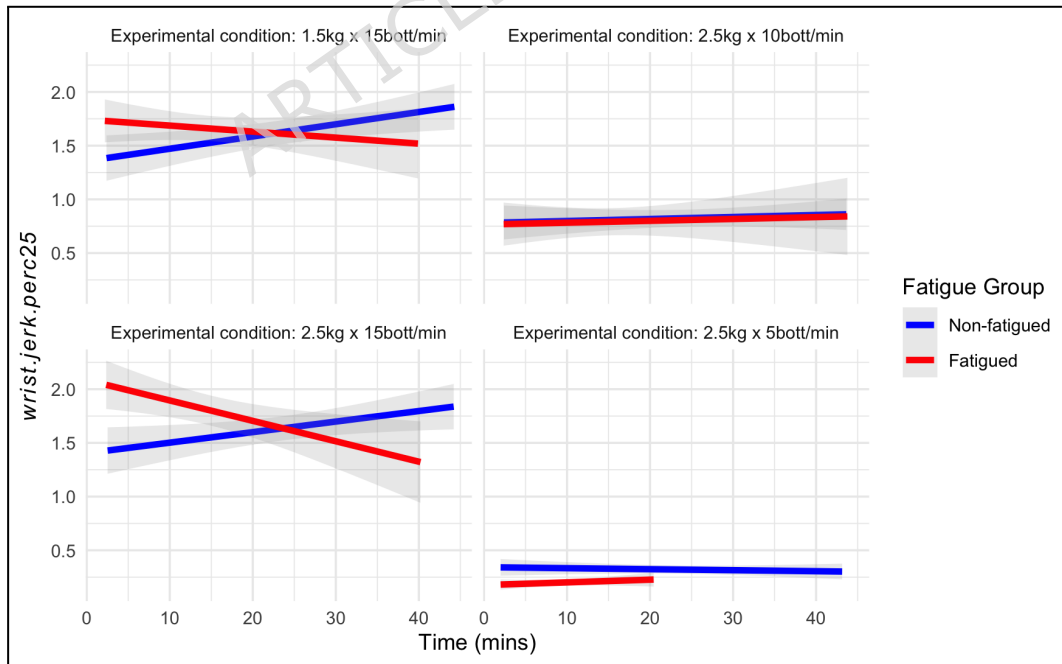


Figure 6. Linear trend lines of *wrist.jerk.perc25* during the BP task, grouped by experimental condition and fatigue group. Shaded regions represent 95% confidence intervals. Each panel represents one experimental condition defined by bottle mass (kg) $\times$ picking rate (bott/min), and compares the overall rate of change in movement smoothness for fatigued (red) vs. non-fatigued (blue) groups.

Residual Check for Both Case Studies

Residual diagnostics indicated an acceptable model fit for both the longitudinal and survival submodels. For the longitudinal component, the residuals-versus-fitted plot showed no major pattern or heteroscedasticity, suggesting that the linear mixed-effects specification was adequate for capturing the sensor signal trajectories (See Figures 7 and 9). For the survival component, the Cox-Snell residual plot showed that the estimated survival function generally aligned with the unit exponential reference line and remained within the 95% pointwise confidence bands of the KM estimate. Slight deviations at lower residual values indicate that the model may slightly underestimate the hazard for early fatigue events (See Figures 8 and 10).

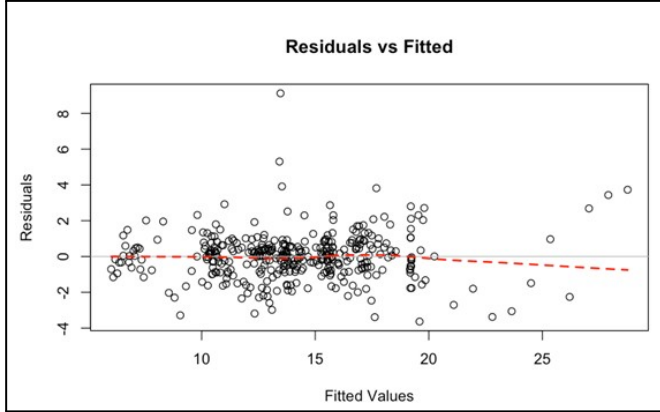


Figure 7. Scatterplot of residuals-versus-fitted values from the longitudinal component of the joint model using *wrist.jerk.mean* feature in Case Study 1.

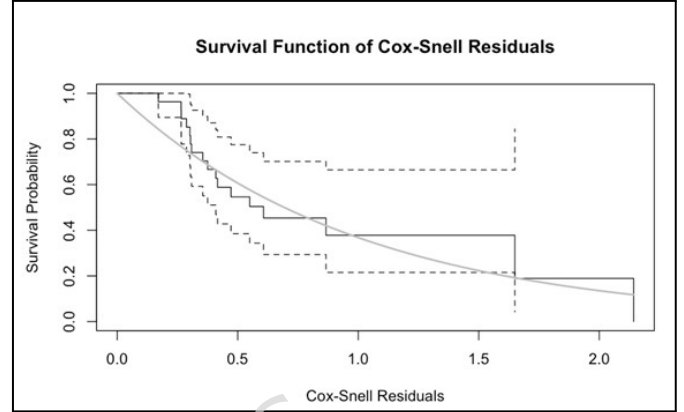


Figure 8. Cox-Snell residual plot for the survival submodel using *wrist.jerk.mean* feature in Case Study 1, with 95% pointwise confidence bands.

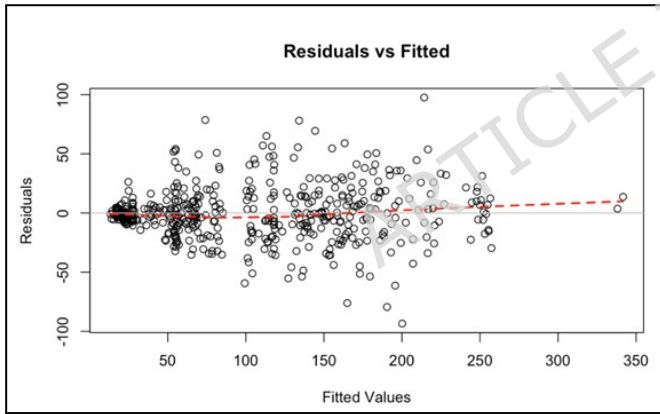


Figure 9. Scatterplot of residuals-versus-fitted values from the longitudinal component of the joint model using *wrist.jerk.perc25* feature in Case Study 2.

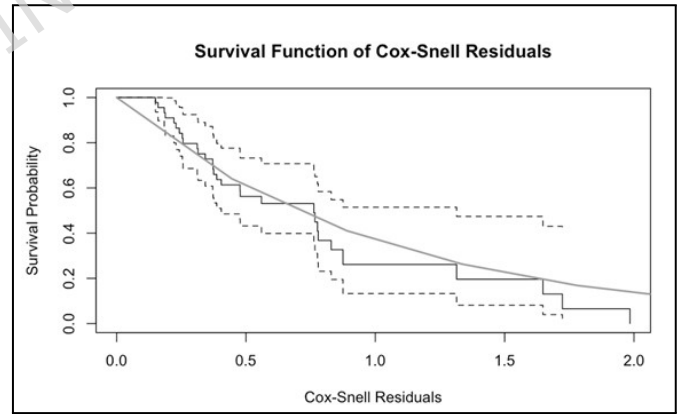


Figure 10. Cox-Snell residual plot for the survival submodel using *wrist.jerk.perc25* feature in Case Study 2, with 95% pointwise confidence bands.

Discussion

In this study, we examined the effectiveness of the JM approach in predicting TTF during simulated manufacturing tasks, using a single sensor's longitudinal trajectory across two case studies involving different work tasks and durations. Across the two independent case studies, the JM approach consistently outperformed baseline models that relied on time-fixed task factors, showing substantial improvements in both discrimination and accuracy of TTF predictions. In line with established performance standards, our joint models achieved a CND% above 70%, indicating good discrimination³⁰ in TTF estimation. Meanwhile, the best MAPE values in two case studies (18-25%) remained within a satisfactory predictive accuracy range, outperforming previously reported benchmarks in fatigue endurance time studies^{20,21}. In particular, both joint models performed well without incorporating time-fixed task factors (i.e., *Task* in Case Study 1 and *Experimental condition* in Case Study 2), probably because the selected sensor features may have accounted for much of the variability associated with these conditions. The

comparable performance between training and test sets—reflected in absolute MAPE differences under 2% and CND% within 5%—indicates minimal overfitting in the best-performing joint models. The best performing joint model in Case Study 1 (MMH and SI tasks) demonstrates slightly higher accuracy of TTF relative to Case Study 2 (BP task), as indicated by the MAPE metrics. This performance difference between case studies may reflect differences in task duration, feature dynamics, and time window configurations. Specifically, the longer duration of Case Study 1 likely allowed for greater within-subject variation in sensor trajectories, while the selected sensor feature and measurement intervals may also have been better aligned with fatigue onset patterns—each of which could enhance the model’s ability to detect fatigue risk over time.

Our results also extend existing fatigue monitoring research by suggesting that jerk-based features, previously identified as indicators of fatigue states in physically demanding tasks^{12,31}, may also serve predictors of TTF. The findings from two case studies indicate that higher *wrist.jerk.mean* values are associated with increased fatigue risk in the simulated MMH or SI tasks, whereas a flat or slight upward trend in *wrist.jerk.perc25* corresponds to delayed fatigue onset in the simulated BP tasks.

The association of increased *wrist.jerk.mean* with the risk of fatigue onset in simulated MMH and SI tasks is consistent with Zhang et al. (2019)¹⁹, who reported elevated mean jerk values in the upper extremities with the transition from unfatigued to fatigued states during masonry lifting tasks. Both masonry lifting and MMH tasks involve repetitive lifting and carrying of heavy loads, activities known to cause neuromuscular fatigue and subsequent changes in movement kinematics. This result also extends the observations from Hostler et al. (2021)³², who found increased acceleration and mean jerk at the wrist during physical work tasks (including walking and lifting) in hot conditions, suggesting fatigue-induced alterations in movement patterns. Generally, elevated jerk magnitudes (i.e., high rates of change of acceleration) imply less controlled movements, increasing mechanical stress and injury risk, whereas smoother movements (lower jerk) are biomechanically preferable for musculoskeletal safety. In one study of construction workers across the workweek, hand tremors increased over the workday, and both hand and shoulder tremors increased over the workweek³³. A study by Yung and Wells (2017)³⁴ found significant differences in action tremor across a range of dynamic fatiguing task conditions, and examined its correlations with other measures in a fatigue test battery. They concluded that peripheral mechanisms are dominant during high intensity activity, whereas fatigue during low intensity activity is more likely attributable to central mechanisms.

In the simulated BP task, the finding that a decreasing slope of the 25th percentile of wrist jerk is associated with a higher hazard of fatigue highlights the utility of distributional metrics in this context and suggests a task-specific biomechanical response to progressive fatigue. For this task, we additionally compared the predictive power of the mean and median of wrist jerk in separate JM analyses. Nonetheless, the 25th percentile emerged as a stronger predictor of TTF in this context. This finding is consistent with previous work by Romijnders et al. (2025). Notably, the significance of a lower percentile in our task, which involves controlled downward placement of bottles, mirrors their finding that the 5th percentile of the peak vertical jerk was the key indicator during the stand-to-sit transition. Whereas Romijnders et al. (2025)³⁵ observed a common fatigue response—an increase in the absolute value of the 5th percentile of jerk as even the smoothest movements become less controlled—our analysis of the temporal slope of the 25th percentile suggests a distinct fatigue response pattern. We interpret the observed negative slope in our simulated task as evidence of a compensatory stiffness-related fatigue strategy. Biomechanically, this may suggest that as the neuromuscular system fatigues from the repetitive BP task, the central nervous system prioritizes stability and accuracy by increasing the co-contraction of agonist and antagonist muscles of the upper extremity. This response is based on the motor control principle in which increasing the impedance of the joint serves to enhance stability and resist perturbations, including internal noise³⁶. Such a stiffening strategy has been observed as a direct response to muscular fatigue, where it acts to preserve performance in multi-joint movements³⁷. By effectively creating a more rigid limb segment, the operator can dampen out micromovements caused by fatigue to maintain bottle placement performance. This compensatory stiffness, while metabolically costly, results in the smoothest portions of the movement becoming even smoother, leading to the observed decline in *wrist.jerk.perc25*. Therefore, the negative slope should not be interpreted as an improvement in motor control, but rather as a quantifiable signature of the body’s adaptive, stiffening strategy or compensatory response to cope with the onset of fatigue. Additionally, this methodological pattern suggests that, in Case Study 2, percentile-based metrics such as *wrist.jerk.perc25* may capture earlier or more nuanced signs of neuromuscular fatigue than measures of central tendency. Given the task-specific experimental setup and the differing fatigue thresholds across studies, this interpretation should be viewed as context-dependent.

Notably, the two sensor features were independently identified as important predictors within their respective task contexts. While we do not claim that different features necessarily indicate different underlying fatigue mechanisms, this finding highlights the importance of task-specific modeling. When validating this approach in occupational settings, practitioners and researchers should carefully consider sensor placement (i.e., body locations) and the statistical form of sensor-derived features, together with relevant time-fixed task covariates, when applying JM for fatigue monitoring across varied tasks. It is also important to note that the observed relationships between sensor features and TTF are associative rather than causal, given the observational nature of the framework.

Overall, these findings and comparisons indicate that the JM approach, leveraging a single sensor trajectory, demonstrates

solid predictive performance for TTF estimation in the simulated work tasks examined. Compared to prior research^{10,12,15,16} that primarily focuses on predicting fatigue occurrence or intensity, the knowledge of TTF may provide more intuitive insights to inform fatigue intervention planning. By continuously updating risk assessments as new sensor data become available, JM has the potential to enable proactive detection of increasing fatigue risks and to support timely interventions (e.g., task rotation, work-rest scheduling) to help reduce injury incidence and maintain productivity in future applied settings.

Limitations and Future Work

This study has three primary limitations. First, the data used in both case studies were collected under controlled laboratory conditions involving simulated work tasks. While these tasks were designed to mimic real manufacturing activities, they may not fully capture the complexities and variability of real-world work environments. Factors such as unexpected interruptions, multitasking demands, ambient noise, interpersonal interactions, and environmental variability (e.g., temperature, lighting) are characteristic of real occupational settings and may influence fatigue development differently. Consequently, the JM approach demonstrates foundational efficacy but is not yet validated for field deployment or real-time operational use. Future research should validate the proposed JM framework in operational environments to assess its real-world applicability and robustness. In addition, future field-based studies should consider evaluating how early fatigue warnings can inform actionable interventions, such as adjustments to work-rest schedules or task allocation.

Second, due to the constraints inherent in human-subject lab experiments, the sample sizes in both case studies were small. Although the cross-validation results of the JM approach suggest minimal overfitting, future research involving larger and more diverse working populations is warranted. Such work would not only help further evaluate the generalizability of the JM framework but also enable the exploration of more flexible modeling structures. For example, expanded datasets would allow meaningful comparisons with machine learning approaches—including tree-based or deep learning survival models adapted for time-varying covariates—which typically require substantially large sample sizes to yield stable estimates³⁸.

Relatedly, we acknowledge that physical fatigue is inherently a multi-joint, systemic phenomenon, and strict reliance on single-feature models risks may oversimplify these complex mechanisms. However, our supplementary multivariate analysis (Tables S3 and S4) confirmed that the univariate approach yielded the best statistical fit for the current datasets, whereas adding candidate features led to model instability. This suggests a trade-off between physiological completeness and statistical parsimony given the available sample size. Future work utilizing larger datasets should further explore multivariate modeling to capture systemic fatigue adaptations without inducing the overfitting observed here. Specifically, exploring physiologically informed dimensionality reduction—such as Principal Component Analysis or weighted feature integration—could allow for the detection of systemic fatigue without substantially increasing model complexity or sacrificing interpretability.

Third, it is important to note that the two case studies were conducted independently, using separate datasets collected at different times with distinct task protocols, sensor placements, and data preprocessing pipelines. Consequently, the available sensor features differed between studies. The primary objective of this work was to demonstrate the feasibility and adaptability of the JM framework across distinct simulated occupational scenarios, rather than to conduct a direct comparison of specific sensor features. As task demands influence sensor placement and feature relevance, our modeling approach focused on task-specific single-sensor inputs. This aligns with our research question, which was to identify sensor features that are predictive of TTF within a given work task. Future research could explore harmonized sensor configurations and feature sets across diverse tasks to enable more direct comparisons and support broader generalizability. We also acknowledge that not all statistical features (e.g., jerk percentiles) were available in both datasets, and future studies could include a broader set of derived features to facilitate systematic comparison across tasks. To ensure that TTF represented a valid functional failure point for each distinct work task, we adopted task-specific fatigue thresholds; future research could explore normalized fatigue metrics (e.g., relative capacity utilization) to facilitate broader cross-task synthesis. Taken together, these extensions will improve the ecological validity and the practical scalability and utility of TTF prediction models in occupational settings.

The current study demonstrates the feasibility of offline training and dynamic predictions using preprocessed sensor data. To translate the JM framework into real-world applications, however, system integration and latency challenges remain to be addressed. We propose a phased implementation approach. The first phase will involve real-time validation in a controlled lab setting to evaluate system latency and software integration. This will be followed by a small-scale pilot study in relevant field environments, such as manufacturing or warehousing, before a full-scale deployment for sustained evaluation and model updates. To minimize latency, model training will be conducted offline using historical sensor and fatigue data, while real-time inference will rely on streaming sensor input processed through a pre-trained model. We plan to explore edge–cloud hybrid processing, where features are extracted on a local device and risk scores computed either locally or via a lightweight API. A sensitivity analysis will also be conducted to optimize prediction update frequency, balancing responsiveness with resource constraints. For integration, the system will use commercial wearable sensors transmitting via Bluetooth to a mobile interface, with fatigue alerts presented through intuitive visual or haptic feedback. These steps aim to support the technical feasibility and operational adoption of JM-based fatigue prediction systems.

When considering the practical application of this JM approach, it is important to reflect on the ethical implications of continuous worker monitoring. Access to individualized sensor data can help identify early signs of physical fatigue and enable timely interventions to protect worker safety. However, robust safeguards are essential to ensure that such data are not misused for productivity surveillance or punitive measures. Practical deployment must include rigorous governance structures, such as implementing source-level anonymization and establishing strict role-based access protocols that separate safety monitoring from managerial performance evaluation. Our modeling approach contributes to this protection by predicting future states from a single, abstracted sensor variable, thereby minimizing the need to process direct task performance data and enhancing worker privacy. A related ethical consideration is fairness. While our modeling approach focuses on individualized prediction, we recognize that certain demographic characteristics—such as gender and fitness level—may influence sensor dynamics and fatigue responses. Our study included a balanced representation of male and female participants, all of whom began the sessions in a rested state. Nonetheless, we intentionally avoided modeling these demographic factors to prevent reinforcing biases or discriminatory practices in occupational health applications. Future research should further explore how to account for individual variability in a way that promotes fairness, inclusivity, and ethical use of wearable technology in the workplace.

Looking forward, beyond real-time individual monitoring, a key opportunity lies in using the knowledge generated by this approach to inform system-level policies. For instance, understanding the principal drivers of fatigue development from individual data can help design evidence-based, generalized work-rest schedules that enhance safety for all workers, including in scenarios without continuous individual monitoring.

Methods

This paper leverages two previously published datasets involving simulated manufacturing tasks as case studies to examine the effectiveness of JM in TTF prediction. Each case study adopted a crossover design, with subjects completing multiple tasks or experimental conditions. This approach increased the number of observations despite modest sample sizes—common in biomechanics and human factors research—and enabled within-subject comparisons to enhance statistical power for detecting task-related effects. Gender balance was maintained across cohorts to support representative sampling and reduce potential bias.

The following subsections outline the key components of data acquisition, preparation, modeling, and evaluation. An overview of the workflow is illustrated in Figure 11.

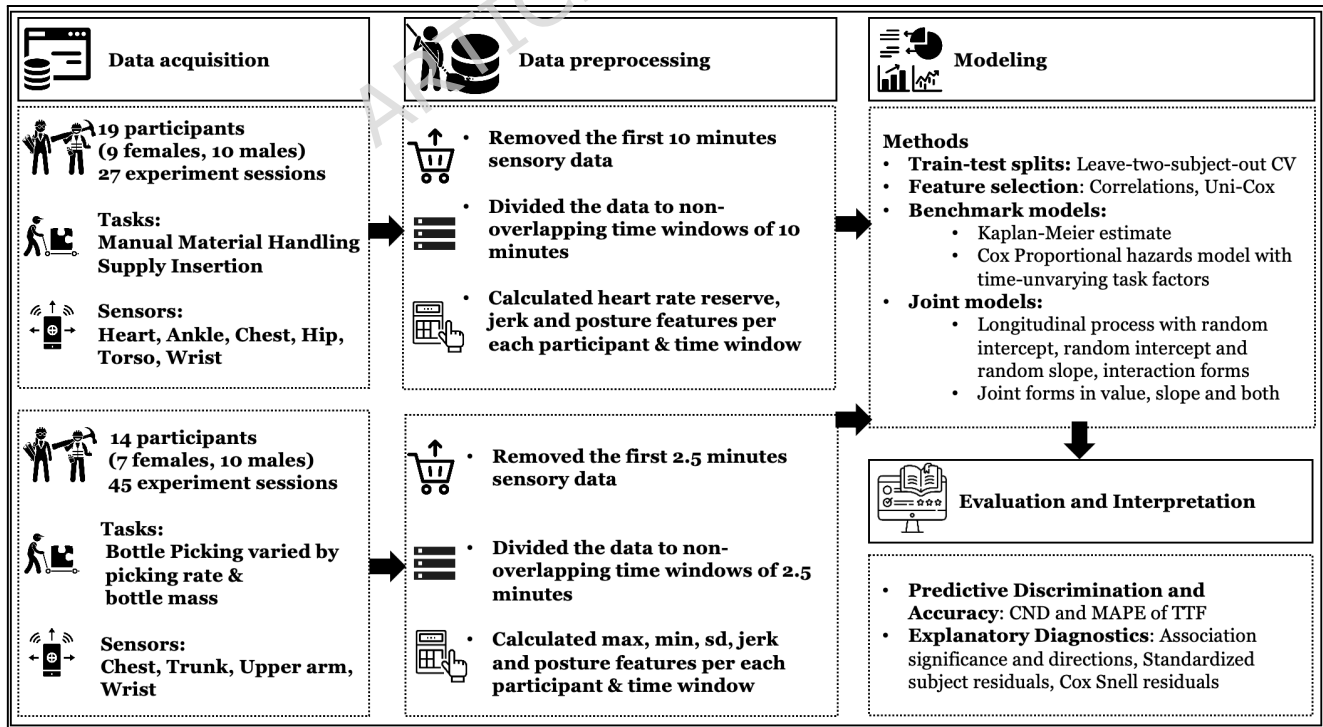


Figure 11. Overview of the methodological workflow

Data Acquisition

Case Study 1: MMH and SI Tasks

The first case study utilizes data from simulated manufacturing work tasks made available in Maman et al.'s paper¹². The experiments involved 19 subjects (nine females with a mean age of 27.75 (standard deviation = 7.94 years), and 11 males with a mean age of 42.18 (standard deviation = 20.3 years). Informed consent was obtained from all subjects prior to the start of the experiment. All study procedures were approved by the Institutional Review Board (IRB) at the University at Buffalo and were conducted in accordance with the relevant ethical guidelines and regulations. Subjects were asked to complete two three-hour sessions—one for the simulated MMH task and one for the SI task. The sessions were conducted in a randomized order on separate days, with no more than one session per day to minimize fatigue carryover effects. No rest breaks were provided during each session to preserve the natural progression of physical fatigue. In the MMH task, the subjects picked, loaded, transported, and palletized weighted containers (10 kg, 18 kg, and 26 kg) from points A to B (Figure 12). In the SI task, subjects walked with supplies, bent forward at the supply box upon reaching the destination, and unscrewed and fastened bolts on the box (Figure 13).



Figure 12. MMH Task

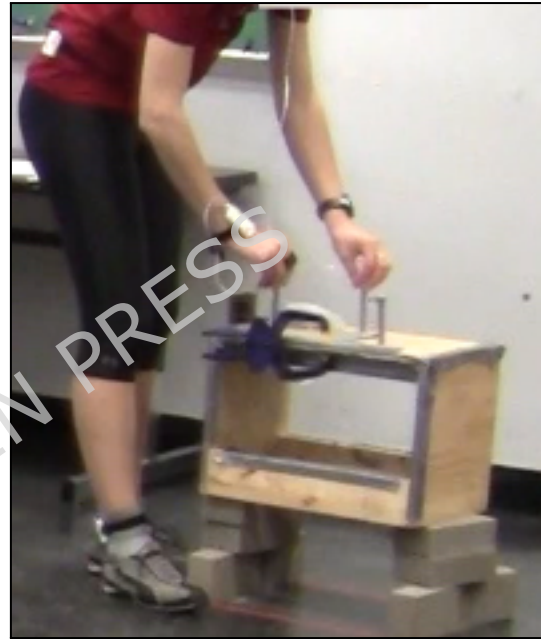


Figure 13. SI Task

Subjects were equipped with four IMUs at the ankle, hip, wrist, and torso, and a heart rate monitor on the chest. Of the 19 subjects, valid data were obtained from 12 individuals for the SI task and from 15 individuals for the MMH task; the rest were excluded due to protocol deviations or sensor malfunctions. To reduce high-frequency noise, a Savitzky-Golay smoothing filter (Polynomial Order: 3, Frame Length: 7) was first applied to the raw data. The data across sensors were synchronized using the data collection software, and then uniformly downsampled to 25 Hz to ensure consistency across all sensors, tasks, and subjects. According to Maman et al.^{12,15}, sensor features were extracted by segmenting the data into non-overlapping 10-minute windows. The 10-minute time window reflects a reasonable task cycle time in occupational settings, such as manufacturing, enabling effective fatigue monitoring without significant loss of information¹⁵. In addition, subjects reported their perceived exertion every 10 minutes using the RPE scale³⁹.

During each measurement window, heart rate reserve (HRR) was derived from the heart rate monitor, while jerk and posture statistics were computed from the IMUs. The first 10-minute data were removed due to the learning effect at the beginning of each task and the minimal level of physical fatigue. The original data consisted of 64 time-varying covariates, including 63 sensor features and one fatigue rating outcome.

Case Study 2: BP task

Our second case study used data from the simulated order picking task from the study by Vahedi et al.⁴⁰ and Kheiri et al.⁸. The experiment involved 14 subjects (seven females with an average age of 26.9 and a standard deviation of 3.9, and seven males with an average age of 26.7 and a standard deviation of 5.5). Prior to initiating the BP task experimental procedures, all individuals provided informed consent. The research protocol received ethical approval from the IRB at the University at

Buffalo and was conducted in accordance with the relevant ethical guidelines and regulations. The sessions were conducted in a randomized order on separate days. Subjects lifted weighted bottles from shoulder height (Figure 14) and placed them into a carton at waist height (Figure 15). Each subject performed this task for up to 45 minutes per session, with four different experimental conditions on separate days to allow fatigue recovery between sessions. The four task conditions vary by bottle mass and picking rate and include: 2.5 kg×15 bott/min, 2.5 kg×10 bott/min, 2.5 kg×5 bott/min, and 1.5 kg×15 bott/min.

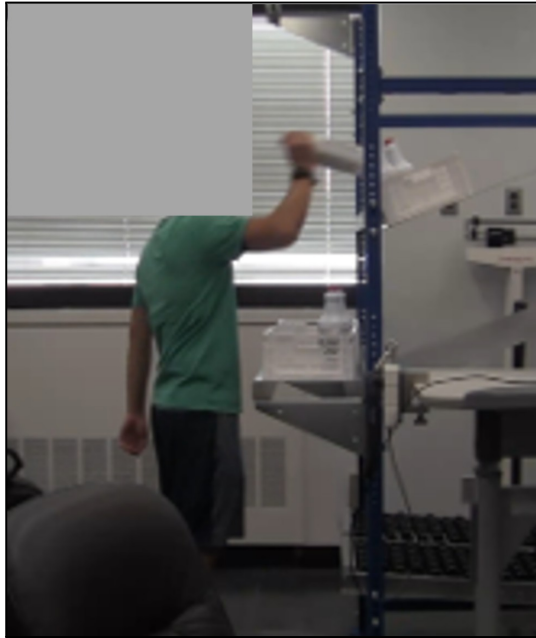


Figure 14. BP Task – Lifting a Bottle



Figure 15. BP Task – Placing a Bottle

Subjects were equipped with three IMUs at the trunk, upper arm, and wrist. Out of 14 subjects, five completed all four sessions, while the remaining subjects completed two or three sessions. Valid data for the BP task were obtained from 45 subject-task sessions; the rest were excluded due to protocol deviations or sensor-related issues. A low-pass Butterworth filter with a 0.2 Hz cutoff frequency was initially applied only to the trunk bending acceleration signal to identify task cycles based on the low frequency bending motions. All statistical features used in the fatigue analysis were subsequently computed from the original, unfiltered sensor data. The data across sensors were synchronized using the data collection software and sampled at 100 Hz to ensure uniformity across sensors, sessions, and subjects. The sensor features were generated by dividing the data into non-overlapping time windows of around 2.5 minutes with the posture and jerk features calculated based on each sensor. The 2.5-minute time window used for the 45-minute task in Case Study 2 is proportional to the 10-minute windows used in the 3-hour tasks of Case Study 1.

The original data contained 55 time-varying covariates including 54 sensor features and one fatigue rating outcome. Although subjects rated shoulder exertion using the Borg CR-10 scale every 5 minutes in the original data, its values were imputed for every 2.5 minutes in this current study to match the corresponding sensor feature measurement frequency. This is also to provide an adequate number of observations to better capture the variation of sensor features for model training and prediction. The initial 2.5-minute records of each session were also excluded to account for the initial learning phase and the negligible level of physical fatigue at the beginning of the task.

Data Preparation

To prepare the outcomes for TTF prediction, three calculated fields were derived from the fatigue rating: the final fatigue indicator, fatigue states per measurement, and TTF. The validity of using RPE as a fatigue measurement in these cases is corroborated by objective physiological data reported in the source studies. In Case Study 1, which utilized data collected by Maman et al. (2020), HRR, jerk, and posture-related metrics were identified as key indicators of developing fatigue during the physically demanding tasks such as MMH and SI¹². For Case Study 2, Vahedi et al. (2023) demonstrated that an increase in subjective fatigue ratings was consistently correlated with a measured decline in relative strength⁴⁰.

With RPE established as a valid measure, specific fatigue thresholds were determined for each case study based on task demands and durations, aligning with established ergonomic research. The 180-min MMH and SI tasks in Case Study 1 involved sustained full-body exertion, making a Borg RPE threshold of 15 on the 6–20 scale appropriate for defining significant

fatigue¹⁵. In contrast, the 45-min BP task, characterized by repetitive upper-limb movements, used a Borg CR-10 RPE threshold of 4 to capture moderate exertion levels typically associated with localized muscle fatigue^{41,42}. These thresholds also resulted in a comparable fatigue incidence across studies: 63% for the MMH/SI tasks and 66.7% for the BP task.

For each case study, we prepared two datasets: a long-format dataset containing repeated sensor measurements per session and a short-format dataset recording the final TTF outcomes. Leave-two-subject-out cross-validation was implemented for both datasets. In each fold, two subjects were held out for testing while models were trained on the remaining subjects. This structure ensures subject-independent evaluation and avoids information leakage during feature selection and model fitting.

Given the high dimensionality of the sensor features measured, rather than conducting an exhaustive search for the most indicative biomarkers, we selected candidate sensor features that showed point-biserial correlations⁴³ of $|r| \geq 0.2$ with time-varying fatigue states in more than half of the leave-two-subject-out cross-validation folds for each case study. The threshold of 0.2 aligns with commonly used benchmarks in behavioral and biomedical research, where it is generally interpreted as exceeding a small effect size and indicative of a meaningful association^{44,45}. To strengthen the robustness of this selection, we also fitted univariate time-dependent Cox proportional hazards models for each sensor feature and retained those that were statistically significant ($p < 0.05$) in over half of the folds. This additional screening accommodates time-varying covariates, provides formal statistical inference⁴⁶ and remains numerically stable given our sample size and censoring level⁴⁷. In the Supplementary Information, Table S1 and S2 present the complete set of variables for the two case studies, respectively.

Baseline Models

To evaluate the effectiveness of JM with a single sensor as a TTF biomarker, we began with a baseline analysis using KM and Cox models with and without time-fixed task-relevant factors. In Case Study 1, we tested the effect of the *Task* factor, (MMH and SI) on TTF, while in Case Study 2, we considered *Experimental condition* (bottle mass and picking rate). The KM method is a non-parametric estimator that does not assume any specific distribution for survival times and accounts for censored observations. While the Cox model is a semi-parametric method which assumes that covariates have a multiplicative effect on the hazard for an event. The corresponding equations are provided in the Supplementary Information.

Joint Models

JM in this study integrates a Cox survival submodel with a linear mixed-effects submodel for longitudinal data, explicitly accounting for endogeneity and measurement noise. To derive a reliable estimate of the true sensor measurement values, three variations of the linear mixed-effects model were evaluated: random intercepts (RI), RIRS⁴⁸, and an interactive form incorporating the effect of task implementation over time. We also piloted a natural spline structure for the longitudinal submodel. Details and justification for this submodel selection are discussed in the Supplementary Information. To fully examine the association between sensor features and fatigue outcomes, in addition to evaluating the biomarker values, the time-dependent slope of the longitudinal trajectory is also considered as a potential indicator of the hazard rate. Accordingly, three forms of parameterizations—value, slope, as well as both value and slope⁴⁹—are considered in JM. The corresponding formulas are provided in the Supplementary Information. To further evaluate the robustness of our univariate JM approach, we conducted targeted multivariate analyses,⁵⁰ in which each bivariate joint model combined the best performing biomarker with one additional candidate sensor feature and assessed whether this could improve overall model fit. These analyses suggested that a well-chosen individual sensor feature can be informative for the present data, supporting the use of univariate JM in this context. Details of these analyses are provided in the Supplementary Information.

JM was first fitted on the training folds of the leave-two-subject-out cross-validation, with dynamic predictions generated using the first six or seven sensor measurements from the longitudinal trajectory in both training and testing folds, respectively. In Case Study 1 (MMH/SI), fatigue ratings and sensor statistics were recorded or calculated every 10 minutes during a 180-minute task, yielding 6–7 longitudinal observations by the selected 60–70 minute follow-up, prior to the earliest fatigue event at 80 minutes. In Case Study 2 (BP), with a shorter total duration of 45 minutes, sensor statistics were calculated around every 2.5 minutes, and RPE ratings (originally recorded every 5 minutes) were interpolated to align with sensor intervals, providing a comparable number of longitudinal observations with Case Study 1. These follow-up time points, representing approximately one-third of the total task duration, were chosen to balance early fatigue prediction, sufficient longitudinal data points for estimation, and comparability across tasks. This proportionate design of measurement intervals across the two studies (180 vs. 45 minutes task duration and 10-minute vs. 2.5-minute measurement intervals) enabled interpretable predictions and methodological consistency across the two studies.

Joint models in this work were implemented in R (version 4.4.1) and executed on a MacBook Pro equipped with an Apple M3 Max processor and 36 GB of RAM. Within the leave-two-subject-out cross-validation framework, each model specification required an average training time of approximately 0.42 seconds per fold and a prediction time of about 0.31 seconds on the test set, with total runtime accumulating across all folds and specifications.

Evaluation Metrics

Leave-two-subjects-out cross-validation was used in both case studies to assess the predictive performance of the baseline and JM methods. This approach facilitates both within-subject and across-subject predictions, enabling a more robust evaluation of predictive performance. To ensure a conservative and timely capture of TTF, we used the lower bounds of the predicted fatigue survival probabilities to calculate TTF in both approaches. Specifically, CND and MAPE were adopted to evaluate model discrimination and predictive accuracy, respectively. These metrics are calculated based on the estimated and actual TTF and their equations are detailed in the Supplementary Information.

Data and Code Availability

The data analyzed in this study originate from previously published works^{12,40}. Additional details on data collection and methodology can be found in these publications. To facilitate reproducibility, the compiled dataset and our code used in this study are available at <https://github.com/Michelle170/Survival>.

References

1. Yung, M. & Wells, R. P. Responsive upper limb and cognitive fatigue measures during light precision work: an 8-hour simulated micro-pipetting study. *Ergonomics* **60**, 940–956 (2017).
2. Lu, L., Megahed, F. M., Sese, R. F. & Cavuoto, L. A. A survey of the prevalence of fatigue, its precursors and individual coping mechanisms among us manufacturing workers. *Appl. ergonomics* **65**, 139–151 (2017).
3. Mohapatra, P. *et al.* Wearable network for multilevel physical fatigue prediction in manufacturing workers. *PNAS nexus* **3**, pgae421 (2024).
4. Yung, M., Manji, R. & Wells, R. P. Exploring the relationship of task performance and physical and cognitive fatigue during a daylong light precision task. *Hum. factors* **59**, 1029–1047 (2017).
5. Kolus, A., Wells, R. & Neumann, P. Production quality and human factors engineering: A systematic review and theoretical framework. *Appl. ergonomics* **73**, 55–89 (2018).
6. Setayesh, A., Grosse, E. H., Glock, C. H. & Neumann, W. P. Determining the source of human-system errors in manual order picking with respect to human factors. *Int. journal production research* **60**, 6350–6372 (2022).
7. Bałys, P., Buła, P., Dziedzic, D. & Uznańska, M. The future of work in automated warehouse from the perspective of the employees. *Contemp. organisation management. Challenges trends* 267–281 (2020).
8. Kheiri, S. K., Vahedi, Z., Sun, H., Megahed, F. M. & Cavuoto, L. A. Human reliability modeling in occupational environments toward a safe and productive operator 4.0. *Int. J. Ind. Ergonomics* **97**, 103479 (2023).
9. Lamooki, S. R. *et al.* A data analytic end-to-end framework for the automated quantification of ergonomic risk factors across multiple tasks using a single wearable sensor. *Appl. ergonomics* **102**, 103732 (2022).
10. Narteni, S., Orani, V., Cambiaso, E., Rucco, M. & Mongelli, M. On the intersection of explainable and reliable ai for physical fatigue prediction. *IEEE Access* **10**, 76243–76260 (2022).
11. Asadi, H., Monfared, S., Athanasiadis, D. I., Stefanidis, D. & Yu, D. Continuous, integrated sensors for predicting fatigue during non-repetitive work: demonstration of technique in the operating room. *Ergonomics* **64**, 1160–1173 (2021).
12. Maman, Z. S. *et al.* A data analytic framework for physical fatigue management using wearable sensors. *Expert. Syst. with Appl.* **155**, 113405 (2020).
13. Baghdadi, A., Cavuoto, L. A. & Crassidis, J. L. Hip and trunk kinematics estimation in gait through kalman filter using imu data at the ankle. *IEEE Sensors J.* **18**, 4253–4260 (2018).
14. Aryal, A., Ghahramani, A. & Becerik-Gerber, B. Monitoring fatigue in construction workers using physiological measurements. *Autom. Constr.* **82**, 154–165 (2017).
15. Maman, Z. S., Yazdi, M. A. A., Cavuoto, L. A. & Megahed, F. M. A data-driven approach to modeling physical fatigue in the workplace using wearable sensors. *Appl. ergonomics* **65**, 515–529 (2017).
16. Hajifar, S. *et al.* A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. *Appl. Ergonomics* **90**, 103262 (2021).
17. Baghdadi, A. *et al.* Monitoring worker fatigue using wearable devices: A case study to detect changes in gait parameters. *J. quality technology* **53**, 47–71 (2021).

18. Lamooki, S. R., Kang, J., Cavuoto, L. A., Megahed, F. M. & Jones-Farmer, L. A. Personalized and nonparametric framework for detecting changes in gait cycles. *IEEE Sensors J.* **21**, 19236–19246 (2021).
19. Zhang, L., Diraneyya, M. M., Ryu, J., Haas, C. T. & Abdel-Rahman, E. M. Jerk as an indicator of physical exertion and fatigue. *Autom. Constr.* **104**, 120–128 (2019).
20. Yi, C. *et al.* Modeling and validation of fatigue and recovery of muscles for manual demolition tasks. *Int. J. Environ. Res. Public Heal.* **19**, 930 (2022).
21. Yi, C. *et al.* Fatigue and recovery of muscles for pulling tasks. *Int. J. Environ. Res. Public Heal.* **19**, 15159 (2022).
22. Iridiastadi, H. & Nussbaum, M. A. Muscle fatigue and endurance during repetitive intermittent static efforts: development of prediction models. *Ergonomics* **49**, 344–360 (2006).
23. Mamaghani, N. K., Shimomura, Y., Iwanaga, K. & Katsuura, T. Muscular performance modeling of the upper limb in static postures. *J. physiological anthropology applied human science* **22**, 149–157 (2003).
24. Papageorgiou, G., Mauff, K., Tomer, A. & Rizopoulos, D. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annu. review statistics its application* **6**, 223–240 (2019).
25. Rizopoulos, D. Introduction to the joint modeling of longitudinal and survival data, with applications in r (2021). https://www.drizopoulos.com/courses/Int/JMwithR_Bern_2021.pdf.
26. Davis-Plourde, K. L. *et al.* Joint models for estimating determinants of cognitive decline in the presence of survival bias. *Epidemiology* **33**, 362–371 (2022).
27. Andrinopoulou, E.-R. *et al.* Dynamic prediction of outcome for patients with severe aortic stenosis: application of joint models for longitudinal and time-to-event data. *BMC cardiovascular disorders* **15**, 1–8 (2015).
28. Field, R. J., Adamson, C., Jhund, P. & Lewsey, J. Joint modelling of longitudinal processes and time-to-event outcomes in heart failure: systematic review and exemplar examining the relationship between serum digoxin levels and mortality. *BMC Med. Res. Methodol.* **23**, 94 (2023).
29. Li, W., Li, L. & Astor, B. C. A comparison of two approaches to dynamic prediction: Joint modeling and landmark modeling. *Stat. medicine* **42**, 2101–2115 (2023).
30. Harrell Jr, F. E., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. medicine* **15**, 361–387 (1996).
31. Zhang, L., Diraneyya, M. M., Ryu, J., Haas, C. T. & Abdel-Rahman, E. Assessment of jerk as a method of physical fatigue detection. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 51739, V01BT02A010 (American Society of Mechanical Engineers, 2018).
32. Hostler, D., Schwob, J., Schlader, Z. J. & Cavuoto, L. Heat stress increases movement jerk during physical exertion. *Front. Physiol.* **12**, 748981 (2021).
33. Yung, M., Bigelow, P. L., Hastings, D. M. & Wells, R. P. Detecting within-and-between-day manifestations of neuromuscular fatigue at work: an exploratory study. *Ergonomics* **57**, 1562–1573 (2014).
34. Yung, M. & Wells, R. P. Documenting the temporal pattern of fatigue development. *IIEE Transactions on Occup. Ergonomics Hum. Factors* **5**, 115–135 (2017).
35. Romijnders, R. *et al.* Association of real life postural transitions kinematics with fatigue in neurodegenerative and immune diseases. *npj Digit. Medicine* **8**, 12 (2025).
36. Hogan, N. The mechanics of multi-joint posture and movement control. *Biol. cybernetics* **52**, 315–331 (1985).
37. Forestier, N. & Nougier, V. The effects of muscular fatigue on the coordination of a multijoint movement in human. *Neurosci. letters* **252**, 187–190 (1998).
38. Shabani, N., Yaseri, M., Alimi, R., Nazemian, F. & Zeraati, H. Dynamic survival analysis via a landmarking-gradient boosting approach and its application to kidney transplant data. *BMC Med. Informatics Decis. Mak.* **25**, 368 (2025).
39. Borg, G. *Borg's perceived exertion and pain scales.* (Human kinetics, 1998).
40. Vahedi, Z. *et al.* The relationship between ratings of perceived exertion (rpe) and relative strength for a fatiguing dynamic upper extremity task: A consideration of multiple cycles and conditions. *J. Occup. Environ. Hyg.* **20**, 136–142 (2023).
41. Williams, N. The borg rating of perceived exertion (rpe) scale. *Occup. medicine* **67**, 404–405 (2017).
42. Bosch, T., Mathiassen, S. E., Visser, B., Looze, M. d. & Dieën, J. v. The effect of work pace on workload, motor variability and fatigue during simulated light assembly work. *Ergonomics* **54**, 154–168 (2011).

43. Revelle, W. *psych: Procedures for Psychological, Psychometric, and Personality Research* (2024). R package version 1.9.12.
44. Cohen, J. Set correlation and contingency tables. *Appl. psychological measurement* **12**, 425–434 (1988).
45. Hemphill, J. F. Interpreting the magnitudes of correlation coefficients. *Am. Psychol.* **58** (2003).
46. Therneau, T., Crowson, C. & Atkinson, E. Using time dependent covariates and time dependent coefficients in the cox model. *Surviv. vignettes* **2**, 1–25 (2017).
47. Peduzzi, P., Concato, J., Feinstein, A. R. & Holford, T. R. Importance of events per independent variable in proportional hazards regression analysis ii. accuracy and precision of regression estimates. *J. clinical epidemiology* **48**, 1503–1510 (1995).
48. Rizopoulos, D., Molenberghs, G. & Lesaffre, E. M. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom. J.* **59**, 1261–1276 (2017).
49. Rizopoulos, D. Jm: An r package for the joint modelling of longitudinal and time-to-event data. *J. statistical software* **35**, 1–33 (2010).
50. Rizopoulos, D. *JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data* (2023). R package version 0.8-3.

Declaration Statement

Prior to the commencement of experimental procedures for both case studies, all subjects provided written informed consent. The research protocols were approved by the IRB at the University at Buffalo: Case Study 1 (MMH and SI tasks), Protocol #: 030-767129; Case Study 2 (BP task), Protocol #: STUDY00003984.

Funding Declaration Statement

This research was funded by the Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health through award R21OH011749.

Author Contributions Statement

L.L.: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - Original Draft, Writing - Review & Editing; Z.S-M.: Data Collection, Data Curation, Writing - Review & Editing; L.C.: Investigation, Data curation, Writing - Review & Editing, Funding acquisition. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.