

From Self-Report to Multimodal Resilience KPIs: A Biomarker-Grounded Evaluation Framework for VR Stress–Recovery Training in Healthcare

Sophia Wanner, USTP

Martin Ernst, USTP

Vanessa Leung, USTP

Abstract: Digital resilience interventions show measurable benefits, yet healthcare evaluation still relies on self-report, limiting comparability and insight into mechanisms and transfer. Virtual reality (VR) can elicit stress, but protocols often omit standardized recovery and multimodal biomarkers. We propose a biomarker-grounded framework for VR stress–recovery training that standardizes within-session windows (baseline, stressor, relaxation, recovery) and derives key performance indicators (KPIs) for reactivity and recovery/adaptation (plus maintenance/transfer where available). In XR2ESILIENCE (Horizon Europe), the protocol is repeated at two embedded biomarker assessments and paired with pre–post questionnaires, enabling feasibility and quality control (QC) testing. Analyses prioritize transparent change estimation, add decision-tree profiling when heterogeneity is expected, and use network methods to test multimodal integration, with privacy-first handling of sensitive modalities.

Keywords: Resilience; virtual reality; digital biomarkers; stress–recovery; multimodal assessment; key performance indicators; quality control; healthcare professionals; network analysis; decision trees

1. INTRODUCTION

Healthcare professionals face persistent cognitive and emotional demands that elevate stress burden and increase risk for burnout and downstream health consequences. Digital interventions to promote psychological resilience have expanded rapidly, and recent synthesis indicates that such interventions can meaningfully improve resilience-related outcomes, while also highlighting heterogeneity in effects and methodological variation in evaluation practices (Ang et al., 2022; Schäfer et al., 2024). However, the prevailing evaluation pattern in resilience training, especially in applied healthcare contexts, still depends primarily on self-reported stress, confidence, or well-being outcomes, even when interventions are designed to affect regulation processes that should also be observable in physiology and behavior (Cortez-Vázquez et al., 2024; Nijland et al., 2021).

This creates two recurring problems. First, self-report-centric evaluation constrains interpretability: it is often unclear whether reported improvements reflect durable regulation skill acquisition, transient expectancy effects, or shifts in appraisal unaccompanied by measurable recovery dynamics. Second, comparability across cohorts and sites is limited, because studies

vary widely in timing, stressor structure, and the presence or absence of recovery phases, making it difficult to compute analogous reactivity and recovery indices (Perez-Valero et al., 2021; Velana et al., 2022). Reviews of VR-supported breathing and stress-management interventions similarly note substantial protocol heterogeneity, underexplored maintenance, and limited multimodal biomarker integration (Cortez-Vázquez et al., 2024; Pancini et al., 2025; Velana et al., 2022).

At the same time, VR is uniquely positioned to support standardization because it can deliver repeatable stressor and relaxation contexts, embed time-locked event markers, and record interaction and movement data unobtrusively. A growing body of work demonstrates that VR-based biobehavioral research can produce meaningful stress responses, while also warning that methodological choices and governance constraints strongly shape validity and interpretability (Shirtcliff et al., 2024). Against this background, the goal of this paper is to specify a pragmatic and defensible evaluation framework for VR stress–recovery resilience training in healthcare that operationalizes resilience improvement through a small set of interpretable, multimodal KPIs, supported by an analysis strategy that scales from basic change estimation to heterogeneity- and system-oriented modelling when feasible. While presented as a generalizable approach, the framework is being implemented and tested within XR2ESILIENCE (Horizon Europe), a multicenter trial of VR-based resilience training for healthcare professionals. In addition to pre–post questionnaires, XR2ESILIENCE includes a standardized stress–recovery biomarker protocol embedded in the first and last training session (Session 1 and Session 8).

2. CONCEPTUAL FOUNDATION: RESILIENCE AS REGULATION ACROSS REACTIVITY, RECOVERY, AND TRANSFER

Resilience is increasingly conceptualized as dynamic adaptation rather than a stable trait, with meaningful individual differences in stress-response and recovery patterns (Bonanno, 2004, 2012). In intervention evaluation, this implies that a single average pre–post effect is rarely sufficient: improvements may occur primarily in recovery dynamics, in reduced reactivity, in increased behavioral flexibility, or in better coupling between self-reported outcomes and objective regulation. A measurement framework that targets only global self-report outcomes risks missing these mechanisms.

A multimodal view treats resilience-relevant functioning as emerging from interacting components, including physiological regulation, cognitive control, behavioral patterns, and subjective appraisal, rather than from a single latent factor. Network perspectives in mental health similarly argue that adaptive functioning is reflected in stable, integrated dependence structures among observed variables, whereas vulnerability may appear as fragmented or maladaptive connectivity patterns (Borsboom, 2017; Borsboom & Cramer, 2013). In the context of VR resilience training, this motivates two complementary evaluation targets: changes in specific KPIs (for example, faster autonomic recovery) and changes in how indicators relate to each other (for example, whether subjective coping becomes more tightly linked to objective regulation markers after training).

3. MEASUREMENT DESIGN: STANDARDIZED WITHIN-SESSION WINDOWS AND REPEATED-SESSION LOGIC

3.1. Time-Locked Within-Session Structure

A central design choice is to anchor all biomarker-derived indices to a common within-session structure that explicitly separates baseline, stress reactivity, active regulation, and recovery. The framework therefore uses four time-locked windows: a pre-stressor baseline, a stressor phase, a relaxation/regulation phase, and a post-relaxation recovery phase. As summarized in Figure 1 (Panel A), phase-wise features are derived within these windows so that reactivity and recovery can be estimated from standardized contrasts. Phase boundaries are implemented via event markers to ensure consistent windowing across sites and across the two biomarker assessment sessions (see Section 4.1). This structure directly addresses a widely noted gap in VR stress-recovery research, namely that post-recovery phases are often missing or not comparable to the initial baseline, limiting recovery-slope analyses and between-session comparability (Cortez-Vázquez et al., 2024; Perez-Valero et al., 2021; Velana et al., 2022).

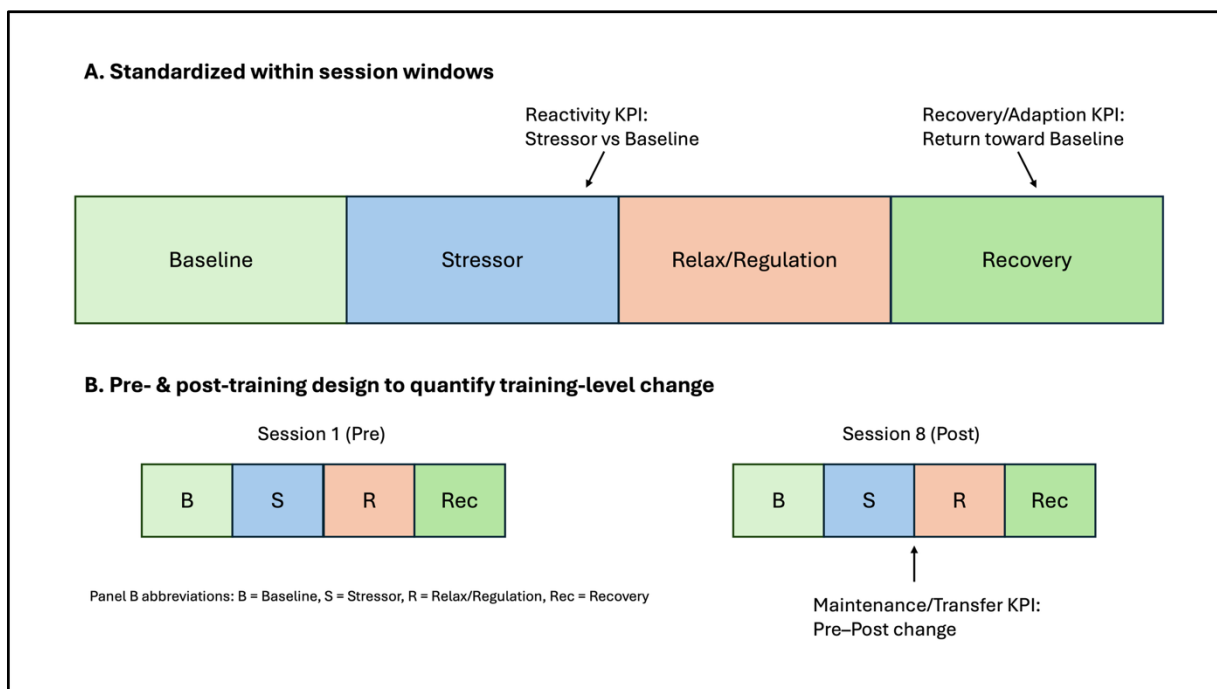


Figure 1: Standardized within-session windows and pre-post assessment design for KPI estimation. Panel A shows the four time-locked windows used for phase-wise feature extraction (baseline, stressor, relax/regulation, recovery), enabling derivation of reactivity (stressor vs. baseline) and recovery/adaptation (return toward baseline). Panel B shows implementation of the same window structure at two biomarker assessment timepoints embedded in the intervention (Session 1 and Session 8), enabling harmonized estimation of training-level change (pre-post differences; additional assessment sessions can be added analogously). Brief in-protocol anchor ratings (not shown) may be time-locked within the biomarker protocol (Section 5.3), while global questionnaire outcomes are assessed separately pre and post. Phase boundaries are time-locked via event markers (see Section 4.1).

3.2. Repeated-Session Logic for Training-Level Change

The same within-session stress–recovery structure is implemented at two standardized biomarker assessment timepoints embedded in the intervention: Session 1 (pre) and Session 8 (post). Figure 1 (Panel B) illustrates how this repeated-window design supports harmonized estimation of training-level change in reactivity and recovery/adaptation. These biomarker timepoints complement separate pre–post questionnaires that capture global self-reported outcomes. Within-session analyses focus on acute regulation dynamics, such as whether physiological signals show a stress-consistent shift during the stressor and a return toward baseline during recovery. Training-level change then captures whether these dynamics improve from pre to post, for example through smaller stress peaks, steeper recovery slopes, or improved alignment between subjective and objective indicators. In XR2ESILIENCE, this repeated-window design enables harmonized KPI estimation across sites at Session 1 and Session 8. This two-timepoint implementation is a pragmatic minimum; additional assessments (e.g., mid-intervention or follow-up) can be added where feasible.

3.3. Modality Selection Under Feasibility Constraints

Modality selection follows a tiered feasibility logic. Core modalities are unobtrusive and broadly available in real-world VR deployments; complementary modalities add mechanistic specificity but depend on additional hardware and stricter quality thresholds; exploratory modalities are included only where privacy and confounding can be managed. Table 1 summarizes candidate modalities by tier, alongside key feasibility constraints and supporting references.

Core modalities prioritize signals that can be captured reliably with minimal additional burden. HR/HRV is included as a core autonomic indicator due to established metrics and interpretability for stress–recovery cycles (Bernardi et al., 2000; Shaffer & Ginsberg, 2017), and slow-paced breathing paradigms embedded in immersive VR contexts have demonstrated feasibility and psychophysiological effects (Blum et al., 2019; Rockstroh et al., 2021). Behavioral markers available by default in VR are also treated as core, particularly head-mounted display (HMD) kinematics and basic interaction telemetry. These streams can be summarized into phase-wise features that may reflect engagement and attentional scanning patterns across the standardized windows, while remaining sensitive to task demands and tracking artefacts (Richer et al., 2024; Ries et al., 2025; Salehi et al., 2024).

Complementary modalities are used where hardware and data quality permit. EDA can add sensitivity to sympathetic arousal and stress responsivity but is more susceptible to contact- and motion-related artefacts in applied deployments and therefore treated as complementary rather than assumed by default (Perez-Valero et al., 2021; Rockstroh et al., 2021; Shirtcliff et al., 2024). Eye tracking can enrich interpretation of attentional control but requires explicit feasibility gates for calibration reliability and spatial accuracy (Schuetz & Fiehler, 2022) and careful metric selection (Skaramagkas et al., 2023). EEG can provide neurophysiological triangulation in VR stress–recovery paradigms (Perez-Valero et al., 2021), but its setup burden and motion artefact sensitivity make it best suited as a complementary modality with strict QC thresholds.

Speech and voice features are treated as exploratory rather than core because the evidence base in VR training contexts is still emerging and privacy and confounding issues are substantial.

Nonetheless, acoustic and prosodic features can reflect physiological stress under certain conditions (Kappen et al., 2024), and synthesis work indicates measurable links between stress and acoustic correlates while underlining variability and methodological constraints (Schewski et al., 2025). Where speech is included, the framework assumes a privacy-first approach that prioritizes feature extraction over retention of raw audio.

Table 1: Candidate modalities for biomarker-grounded evaluation in VR stress–recovery training, grouped by feasibility tier with example constraints and key references.

Modality / signal	Tier	Primary contribution	Feasibility concerns (examples)	Key references
HR / HRV (HR; RMSSD/SDNN)	Core	Autonomic reactivity + recovery dynamics	Beat quality; ectopic/artefact handling; min window length; harmonized HRV settings	Bernardi et al., 2000; Blum et al., 2019; Rockstroh et al., 2021; Shaffer & Ginsberg, 2017
HMD kinematics + telemetry (velocity; movement variability)	Core	Unobtrusive behavioral signatures in VR	Tracking coverage; dropout flags; task vs stress motion separation	Richer et al., 2024; Ries et al., 2025; Salehi et al., 2024
EDA (SCL; phasic peaks)	Complementary	Sympathetic arousal / stress responsivity	Sensor contact; motion/sweat artefacts; min window length; consistent peak rules	Perez-Valero et al., 2021; Rockstroh et al., 2021; Shirtcliff et al., 2024
Eye tracking (fixations; saccades; pupil where supported)	Complementary	Attentional/oculomotor markers	Calibration reliability; drift; validity/precision threshold; illumination notes (pupil)	Schuetz & Fiehler, 2022; Skaramagkas et al., 2023
EEG (bandpower indices)	Complementary	Neurophysiological triangulation (where feasible)	Setup burden; motion artefacts; channel QC; min usable segments; harmonized preprocessing	Perez-Valero et al., 2021; Shirtcliff et al., 2024
Speech / voice features (acoustic/prosodic; features privacy-first)	Exploratory	Stress correlates under standardized elicitation	Opt-in consent; feature-only storage; elicitation standardization; speech–autonomic confounding	Kappen et al., 2024; Schewski et al., 2025; Rantanen et al., 2021

4. HARMONIZED DATA PIPELINE: EVENT MARKERS, QUALITY CONTROL, AND DROP RULES

4.1. Synchronization and Event Marking

Multimodal inference requires precise temporal alignment across streams. The within-session structure is therefore instrumented with required event markers for phase transitions and additional markers for self-report anchor points. These markers support phase-wise feature extraction and enable comparability across sessions and sites.

4.2. Preprocessing and Feature Extraction Principles

Preprocessing must explicitly address artefacts and sensor-specific noise sources, using modality-specific quality gates and harmonized feature definitions (Table 1). For HR/HRV, this includes beat detection quality, ectopic handling, and consistent HRV calculation decisions (Shaffer & Ginsberg, 2017). For eye tracking, calibration quality and drift must be tracked and used to define minimum usable segments (Schuetz & Fiehler, 2022). For VR-derived behavioral features (e.g., HMD kinematics and interaction telemetry), preprocessing should distinguish task-driven behavior from stress- or regulation-related change, and from technical tracking artefacts. Complementary modalities (e.g., EDA, EEG) require additional artefact handling and harmonized preprocessing choices consistent with their feasibility tier.

4.3. Quality Control and Drop Rules

Applied VR studies frequently encounter missingness and data loss. Quality control therefore requires predefined minimum requirements and conservative but explicit drop rules, protecting validity while avoiding post-hoc flexibility. This is particularly important for cross-site comparability.

5. KPI DERIVATION: REACTIVITY, RECOVERY/ADAPTATION, AND TRANSFER

5.1. KPI Definition and Fit-for-Purpose Logic

KPI construction follows an explicit fit-for-purpose approach. KPIs are theory-guided, empirically tractable indicators that translate multimodal signals into interpretable outcomes for evaluation and mechanistic inference. This aligns with established biomarker terminology principles that emphasize contextual validation rather than assuming that sensor features are direct proxies for complex states (FDA-NIH Biomarker Working Group, 2016).

5.2. KPI Categories

Reactivity KPIs capture the magnitude and immediacy of stress responses during the stressor window relative to baseline, for example peak EDA response or HRV suppression indices. These KPI contrasts map directly onto the within-session windows illustrated in Figure 1. Recovery and

adaptation KPIs capture the speed and completeness of return toward baseline during the recovery window, for example HRV recovery slopes and post-relaxation normalization of EDA. Transfer and maintenance KPIs capture whether improvements persist over time and generalize beyond the assessment context, operationalized as pre–post change in reactivity and recovery indices, with maintenance quantified at follow-up where available and transfer evaluated via associations with broader outcomes.

5.3. Minimal-Burden Self-Report Anchors for Interpretability

The framework links objective KPIs to minimal-burden subjective checkpoints within the biomarker protocol rather than extensive within-session questionnaires. Time-locked perceived-stress checks provide a pragmatic manipulation/interpretability check that the stressor elicited a subjective response and that relaxation shifted appraisal, without compromising comparability of physiological windows. Where feasible, brief structured affect reporting can be used sparingly to contextualize biomarker changes (*The Geneva Emotion Wheel - Swiss Center For Affective Sciences - UNIGE*, 2016). These in-protocol anchor ratings are distinct from the broader pre–post questionnaire battery, which captures global self-reported outcomes and is analyzed separately (e.g., for responder profiling and network models).

6. STATISTICAL MODELLING STRATEGY: FROM TRANSPARENT CHANGE ESTIMATION TO HETEROGENEITY AND NETWORKS

6.1. Variable-Centered Change Models as the Default

The analysis plan begins with variable-centered models to estimate within-session and training-level KPI change using pre-specified contrasts between windows and the two biomarker assessment sessions (Session 1 and Session 8), alongside sensitivity checks for plausible confounds and signal-quality thresholds. This layer prioritizes interpretability and robustness and provides a baseline against which more complex models can be judged.

6.2. Heterogeneity Modelling and Person-Centered Subgroups

Resilience research emphasizes substantial individual differences in response patterns and adaptation, and analogous heterogeneity is expected in KPI change profiles (Bonanno, 2004, 2012). When only two biomarker assessment timepoints are available, heterogeneity can be operationalized as responder versus non-responder patterns or as distinct configurations of reactivity and recovery change across modalities. Pragmatic subgroup identification can be performed using tree-based approaches that yield interpretable decision rules for such response patterns (Banerjee et al., 2019; Breiman et al., 2017). Tree models can support practical evaluation questions, but they require conservative complexity constraints and appropriate validation to avoid spurious findings. Figure 2 illustrates how biomarker KPI contrasts between Session 1 and Session 8 (e.g., HRV recovery; HMD movement variability) can be combined with global questionnaire change (pre–post subjective stress) to derive simple, interpretable

responder profiles. Thresholds are shown for illustration; in practice, cutoffs should be pre-specified where possible and validated (e.g., via cross-validation or bootstrapping).

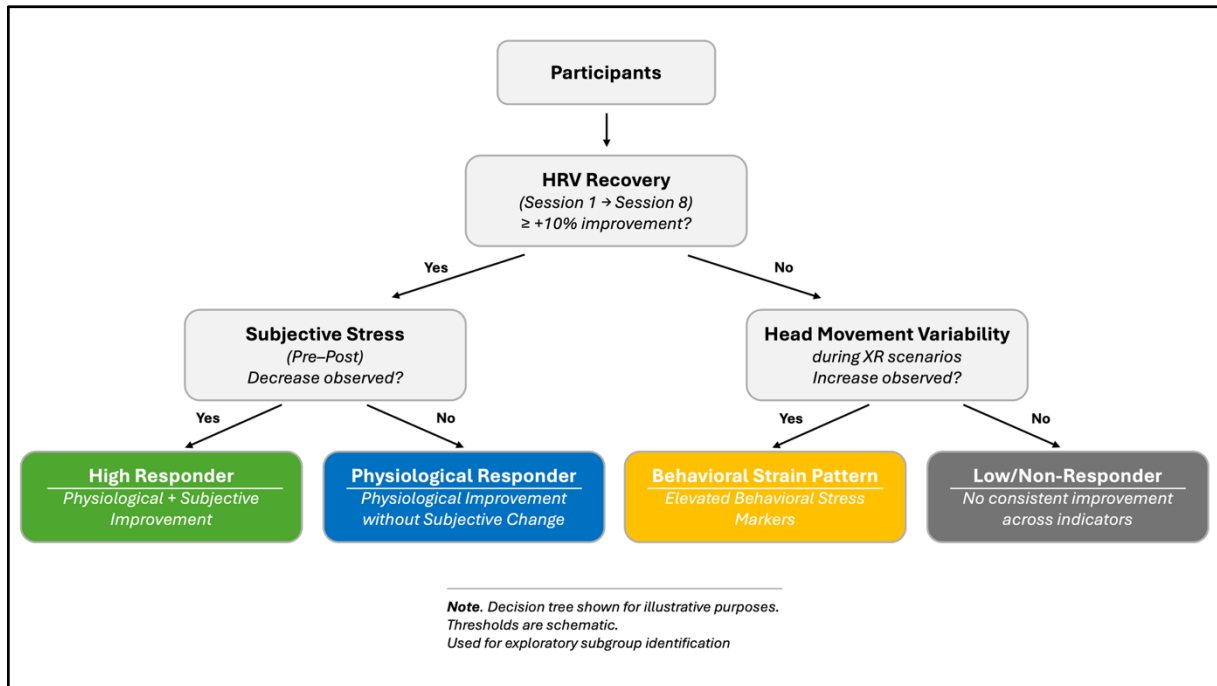


Figure 2: Illustrative rule-based responder profiling using multimodal resilience indicators. Biomarker KPI contrasts between Session 1 and Session 8 (e.g., HRV recovery change; head-movement variability change during XR scenarios) are combined with global questionnaire change in perceived stress (pre-post) to yield example profiles (High Responder, Physiological Responder, Behavioral Strain Pattern, Low/Non-Responder). Thresholds and directions are illustrative; applied analyses should pre-specify and validate decision rules (e.g., cross-validation/bootstrapping).

6.3. Network-Based Multimodal Integration as a System-Level Outcome

A third modelling layer treats resilience as a multimodal system and examines whether training changes the structure of relationships among indicators. Regularized partial correlation networks provide a principled way to model unique associations among many indicators while limiting overfitting (Epskamp et al., 2018; Epskamp & Fried, 2018). Figure 3 provides a schematic example of the indicator networks considered here, where nodes represent subjective and objective indicators and edges represent regularized partial correlations. In this framework, subjective nodes can be operationalized as global questionnaire outcomes (e.g., perceived stress), aligned with the pre and post assessment timepoints. Because centrality metrics can be unstable and conceptually ambiguous, interpretation focuses on robust edges and global properties unless stability diagnostics support stronger claims (Bringmann et al., 2019). To test intervention-related reorganization rather than merely visualizing networks, permutation-based network comparison approaches can evaluate differences in overall structure and global strength across time and between groups (van Borkulo et al., 2023). This supports the mechanistic evaluation question of whether self-reported outcomes become more coherently connected to objective regulation markers after training.

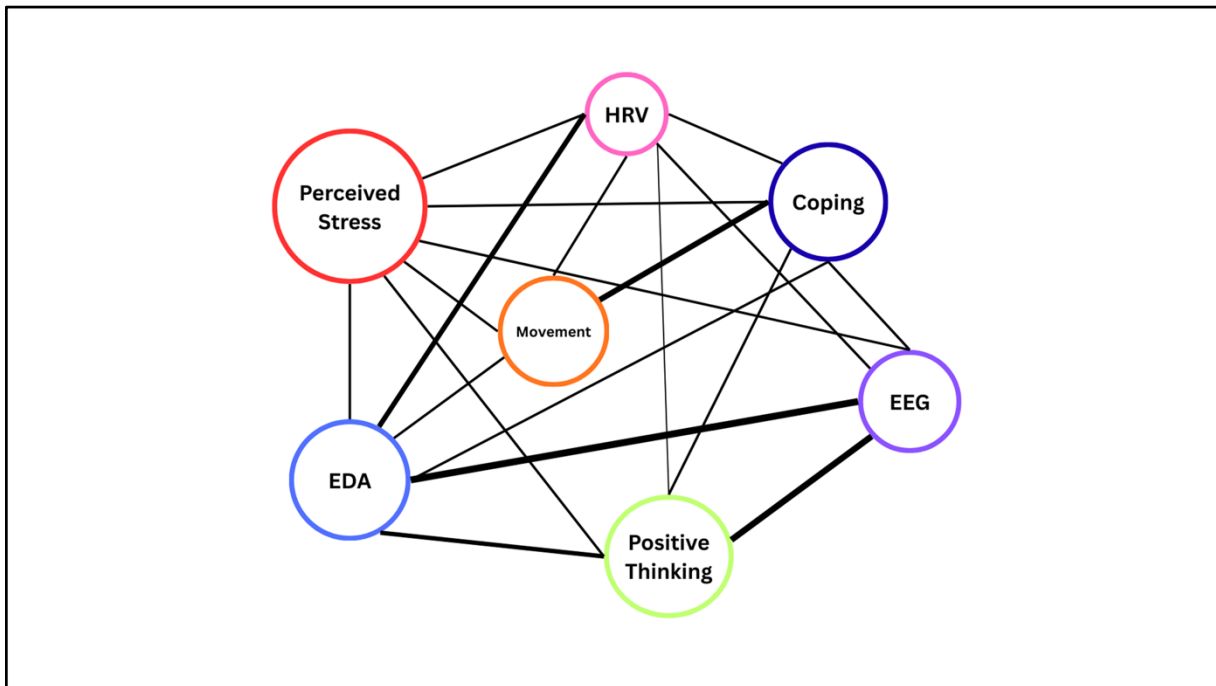


Figure 3: Conceptual schematic of network-based multimodal integration. Nodes represent subjective and objective variables used in the framework, including global questionnaire outcomes (e.g., perceived stress) and biomarker-derived indicators (e.g., HRV, HMD-derived movement features, and complementary signals such as EDA and EEG). Edges represent regularized partial correlations (thicker edges indicate stronger conditional associations). In the proposed analysis plan, network structure is compared between pre and post assessment timepoints (aligned with Session 1 and Session 8) to test whether self-reported outcomes (i.e. Coping, Positive Thinking) become more coherently connected to objective regulation markers after training.

7. ETHICS, PRIVACY, AND FEASIBILITY IN HEALTHCARE DEPLOYMENTS

7.1. Data Protection and Security Baselines

Multimodal biomarker evaluation in healthcare training introduces privacy and governance challenges that must be addressed explicitly. Data protection requirements and information security controls are treated as baseline constraints for the entire pipeline, including remote capture and transfer (ISO/IEC 27001:2022(En), *Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements*, n.d.)

7.2. Privacy-First Handling of Sensitive Modalities

Speech and voice data illustrate why privacy-first handling is essential. Raw audio is highly identifying, so the framework assumes an approach that privileges extracting non-identifying features, limits retention of raw files, and relies on explicit opt-in consent with clear usage boundaries. Interpretation must also consider that speech itself can influence autonomic activity, making confounding plausible (Rantanen et al., 2021).

7.3. Practical Feasibility and Measurement Constraints

Remote deployments increase the importance of automated metadata capture and quality control because variation in sensor fit, environmental noise, and participant routines can substantially affect derived features. VR-specific measurement constraints, such as eye-tracker calibration reliability, are treated as feasibility gates rather than assumed capabilities (Schuetz & Fiehler, 2022). Accessibility considerations also need to be treated as design requirements to avoid systematically excluding subgroups of the workforce (Cortez-Vázquez et al., 2024). XR2ESILIENCE serves as the primary deployment context, which makes these requirements, including automated metadata capture, standardized quality control, and modality-specific feasibility gates (e.g., for eye tracking and EEG), central design constraints rather than optional enhancements, particularly for the two embedded biomarker assessment sessions.

8. DISCUSSION AND CONCLUSION

This paper proposes a pragmatic path for evaluating VR-based resilience training in healthcare that responds directly to limitations repeatedly noted in current evidence, including heavy reliance on self-report, scarce multimodal biomarker integration, missing or non-comparable recovery phases, and protocol heterogeneity that impairs comparability (Cortez-Vázquez et al., 2024; Perez-Valero et al., 2021; Velana et al., 2022). By standardizing within-session windows and repeating the stress–recovery protocol at two embedded biomarker assessment sessions (Session 1 and Session 8), the framework supports harmonized estimation of training-level change. Combined with separate pre–post questionnaires and KPIs that represent reactivity, recovery/adaptation, and transfer, this design enables a clearer separation between immediate elicitation, regulation practice, and durable change. An important next step is empirical evaluation of KPI sensitivity, robustness under missingness, and cross-site comparability in XR2ESILIENCE, which provides an applied healthcare setting for testing the framework’s assumptions.

The analytic strategy is intentionally layered. It begins with transparent KPI change estimation, which is necessary for interpretability and stakeholder trust, and then escalates to heterogeneity and network-based analyses only where the evaluation questions and data quality justify them. This aligns with methodological guidance in psychological network research, which stresses quantifying uncertainty and robustness rather than over-interpreting network visualizations (Epskamp et al., 2018), and it is compatible with expectations for explainability and responsible deployment in healthcare-oriented AI governance discussions (European Commission, 2019; Fehr et al., 2024).

Several limitations remain inherent to biomarker-grounded evaluation. Physiological indicators are sensitive to confounds such as sleep, caffeine, medication, and contextual factors, and they cannot be interpreted as diagnostic markers of mental health states. VR-induced stressors also raise questions about ecological validity and generalizability to real workplace stressors, even when paradigms are well controlled (Shirtcliff et al., 2024). For these reasons, the framework treats biomarker outputs as regulation-relevant signals that gain meaning through standardized context, repeated measurement, and triangulation with minimal subjective checks and longer-term outcomes.

Overall, a biomarker-grounded approach to evaluating VR resilience training in healthcare is feasible if it is designed around standardized stress–recovery windows, repeated-session logic, and a small set of interpretable KPIs. The framework presented here strengthens comparability across settings, supports mechanistic interpretation of regulation change, and creates an extensible base for future personalization, while keeping privacy and feasibility constraints central from the outset.

References

- Ang, W. H. D., Chew, H. S. J., Dong, J., Yi, H., Mahendren, R., & Lau, Y. (2022). Digital training for building resilience: Systematic review, meta-analysis, and meta-regression. *Stress and Health*, 38(5), 848–869. <https://doi.org/10.1002/smi.3154>
- Banerjee, M., Reynolds, E., Andersson, H. B., & Nallamotheu, B. K. (2019). Tree-Based Analysis. *Circulation: Cardiovascular Quality and Outcomes*, 12(5), e004879. <https://doi.org/10.1161/CIRCOUTCOMES.118.004879>
- Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., & Sleight, P. (2000). Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *JACC*, 35(6), 1462–1469. [https://doi.org/10.1016/S0735-1097\(00\)00595-7](https://doi.org/10.1016/S0735-1097(00)00595-7)
- Blum, J., Rockstroh, C., & Göritz, A. S. (2019). Heart Rate Variability Biofeedback Based on Slow-Paced Breathing With Immersive Virtual Reality Nature Scenery. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02172>
- Bonanno, G. A. (2004). Loss, Trauma, and Human Resilience: Have We Underestimated the Human Capacity to Thrive After Extremely Aversive Events? *American Psychologist*, 59(1), 20–28. <https://doi.org/10.1037/0003-066X.59.1.20>
- Bonanno, G. A. (2012). Uses and abuses of the resilience construct: Loss, trauma, and health-related adversities. *Social Science & Medicine*, 74(5), 753–756. <https://doi.org/10.1016/j.socscimed.2011.11.022>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D., & Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, 9(Volume 9, 2013), 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and Regression Trees*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Bringmann, L. F., Elmer, T., Epskamp, S., Krause, R. W., Schoch, D., Wichers, M., Wigman, J. T. W., & Snippe, E. (2019). What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology*, 128(8), 892–903. <https://doi.org/10.1037/abn0000446>
- Cortez-Vázquez, G., Adriaanse, M., Burchell, G. L., Ostelo, R., Panayiotou, G., & Vlemincx, E. (2024). Virtual Reality Breathing Interventions for Mental Health: A Systematic Review and

- Meta-Analysis of Randomized Controlled Trials. *Applied Psychophysiology and Biofeedback*, 49(1), 1–21. <https://doi.org/10.1007/s10484-023-09611-4>
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), 195–212. <https://doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods*, 23(4), 617–634. <https://doi.org/10.1037/met0000167>
- European Commission. (2019). *Ethics Guidelines for Trustworthy AI* [Report]. European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419
- FDA-NIH Biomarker Working Group. (2016). *BEST (Biomarkers, EndpointS, and other Tools) Resource*. Food and Drug Administration (US). <http://www.ncbi.nlm.nih.gov/books/NBK326791/>
- Fehr, J., Citro, B., Malpani, R., Lippert, C., & Madai, V. I. (2024). A trustworthy AI reality-check: The lack of transparency of artificial intelligence products in healthcare. *Frontiers in Digital Health*, 6. <https://doi.org/10.3389/fdgth.2024.1267290>
- ISO/IEC 27001:2022(en), *Information security, cybersecurity and privacy protection—Information security management systems—Requirements*. (n.d.). Retrieved October 28, 2025, from <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:27001:ed-3:v1:en>
- Kappen, M., Vanhollebeke, G., Van Der Donckt, J., Van Hoecke, S., & Vanderhasselt, M.-A. (2024). Acoustic and prosodic speech features reflect physiological stress but not isolated negative affect: A multi-paradigm study on psychosocial stressors. *Scientific Reports*, 14(1), 5515. <https://doi.org/10.1038/s41598-024-55550-3>
- Nijland, J. W. H. M., Veling, W., Lestestuiver, B. P., & Van Driel, C. M. G. (2021). Virtual Reality Relaxation for Reducing Perceived Stress of Intensive Care Nurses During the COVID-19 Pandemic. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.706527>
- Pancini, E., Di Natale, A. F., & Villani, D. (2025). Breathing in virtual reality for promoting mental health: A scoping review. *Virtual Reality*, 29(1), 29. <https://doi.org/10.1007/s10055-024-01096-8>
- Perez-Valero, E., Vaquero-Blasco, M. A., Lopez-Gordo, M. A., & Morillas, C. (2021). Quantitative Assessment of Stress Through EEG During a Virtual Reality Stress-Relax Session. *Frontiers in Computational Neuroscience*, 15, 684423. <https://doi.org/10.3389/fncom.2021.684423>
- Rantanen, A., Laukka, S., Siipo, A., Tiinanen, S., Tarvainen, M. P., Kortelainen, J., Lehtihalmes, M., & Seppänen, T. (2021). Verbalization has regulatory influences on autonomic activity during recall of unpleasant experience. *Speech Communication*, 126, 1–8. <https://doi.org/10.1016/j.specom.2020.11.002>
- Richer, R., Koch, V., Abel, L., Hauck, F., Kurz, M., Ringgold, V., Müller, V., Küderle, A., Schindler-Gmelch, L., Eskofier, B. M., & Rohleder, N. (2024). Machine learning-based detection of acute psychosocial stress from body posture and movements. *Scientific Reports*, 14(1), 8251. <https://doi.org/10.1038/s41598-024-59043-1>

- Ries, A. J., Callahan-Flintoft, C., Madison, A., Dankovich, L., & Touryan, J. (2025). Decoding target discriminability and time pressure using eye and head movement features in a foraging search task. *Cognitive Research: Principles and Implications*, 10(1), 53. <https://doi.org/10.1186/s41235-025-00657-y>
- Rockstroh, C., Blum, J., & Göritz, A. S. (2021). A mobile VR-based respiratory biofeedback game to foster diaphragmatic breathing. *Virtual Reality*, 25(2), 539–552. <https://doi.org/10.1007/s10055-020-00471-5>
- Salehi, M., Javadpour, N., Beisner, B., Sanaei, M., & Gilbert, S. B. (2024). Cybersickness Detection Through Head Movement Patterns: A Promising Approach. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI* (pp. 239–254). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60611-3_18
- Schäfer, S. K., von Boros, L., Schaubruch, L. M., Kunzler, A. M., Lindner, S., Koehler, F., Werner, T., Zappalà, F., Helmreich, I., Wessa, M., Lieb, K., & Tüscher, O. (2024). Digital interventions to promote psychological resilience: A systematic review and meta-analysis. *Npj Digital Medicine*, 7(1), 30. <https://doi.org/10.1038/s41746-024-01017-8>
- Schewski, L., Doss, M. M., Beldi, G., & Keller, S. (2025). Measuring negative emotions and stress through acoustic correlates in speech: A systematic review. *PLOS ONE*, 20(7), e0328833. <https://doi.org/10.1371/journal.pone.0328833>
- Schuetz, I., & Fiehler, K. (2022). Eye Tracking in Virtual Reality: Vive Pro Eye Spatial Accuracy, Precision, and Calibration Reliability. *Journal of Eye Movement Research*, 15(3), 10.16910/jemr.15.3.3. <https://doi.org/10.16910/jemr.15.3.3>
- Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5. <https://doi.org/10.3389/fpubh.2017.00258>
- Shirtcliff, E. A., Finseth, T. T., Winer, E. H., Glahn, D. C., Conrady, R. A., & Drury, S. S. (2024). Virtual stressors with real impact: What virtual reality-based biobehavioral research can teach us about typical and atypical stress responsivity. *Translational Psychiatry*, 14(1), 441. <https://doi.org/10.1038/s41398-024-03129-x>
- Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of Eye Tracking Metrics Involved in Emotional and Cognitive Processes. *IEEE Reviews in Biomedical Engineering*, 16, 260–277. <https://doi.org/10.1109/RBME.2021.3066072>
- The Geneva Emotion Wheel—Swiss Center For Affective Sciences—UNIGE*. (2016, June 29). <https://www.unige.ch/cisa/gew/>
- van Borkulo, C. D., van Bork, R., Boschloo, L., Kossakowski, J. J., Tio, P., Schoevers, R. A., Borsboom, D., & Waldorp, L. J. (2023). Comparing network structures on three aspects: A permutation test. *Psychological Methods*, 28(6), 1273–1285. <https://doi.org/10.1037/met0000476>
- Velana, M., Sobieraj, S., Digutsch, J., & Rinkenauer, G. (2022). The Advances of Immersive Virtual Reality Interventions for the Enhancement of Stress Management and Relaxation among Healthy Adults: A Systematic Review. *Applied Sciences*, 12(14), 7309. <https://doi.org/10.3390/app12147309>