A Quantitative Study on Alert Fatigue in a Novel Experimental Design

Julie Harrison, Zack Tidler, Terri Dunbar, David Grimm, Shiwen Zhou & Stanley Joseph
Human Factors & Ergonomics Society, Georgia Tech Student Chapter

Why investigate alert fatigue?

- Alert fatigue co-occurs with **high volumes of alerts**, especially when they are irrelevant to current workflows or are redundant (Ancker et al., 2017)
- Alert fatigue is generally measured in **number of overrides or other behavioral metrics**, while only select subjective measures are reported in the literature (Ashrafi et al., 2017; Cho et al., 2016; McGreevy, et al., 2019)
- The dominant paradigm for investigating alert fatigue is typically **limited to archival analysis** of EHR data (e.g., Elias et al., 2019)
- Archival analyses may be subject to **limited generalizability** and suffer from a **lack of experimental control**

The present study manipulates variables shown to affect alert fatigue to **understand the strength of their effects** in a novel experimental design.

- **Hypothesis 1:** High alert redundancy will result in higher levels of subjective alert fatigue, but lower levels of workload.
- **Hypothesis 2:** Subjective alert fatigue and workload will be lower as a result of alerting within operator workflow.

Method/Experimental Task

- 201 participants collected from a university population
- Alphabetization task required participants to sort a list of 100 names, first from A to Z, and then again from Z to A
- During task performance, participants were sent to a screen which showed an alert. Participants were asked to accept an alert the first time it was presented, and dismiss the alert on subsequent presentations

- Independent variables:
  - Alert Redundancy (low, medium, high)
  - Point of Interruption (drag, drop)
  - Consistency (variable interval, fixed interval)
- Dependent variables
  - Accuracy
  - Dwell time
  - Alert fatigue (Cho et al., 2016; Ashrafi, Mehri, & Nehrir, 2017)
  - Workload: NASA TLX (Hart & Staveland, 1988)
- Other self-report measures
  - Situation awareness
  - Participant strategy
  - Emotional response: SPANE (Diener et al., 2010)

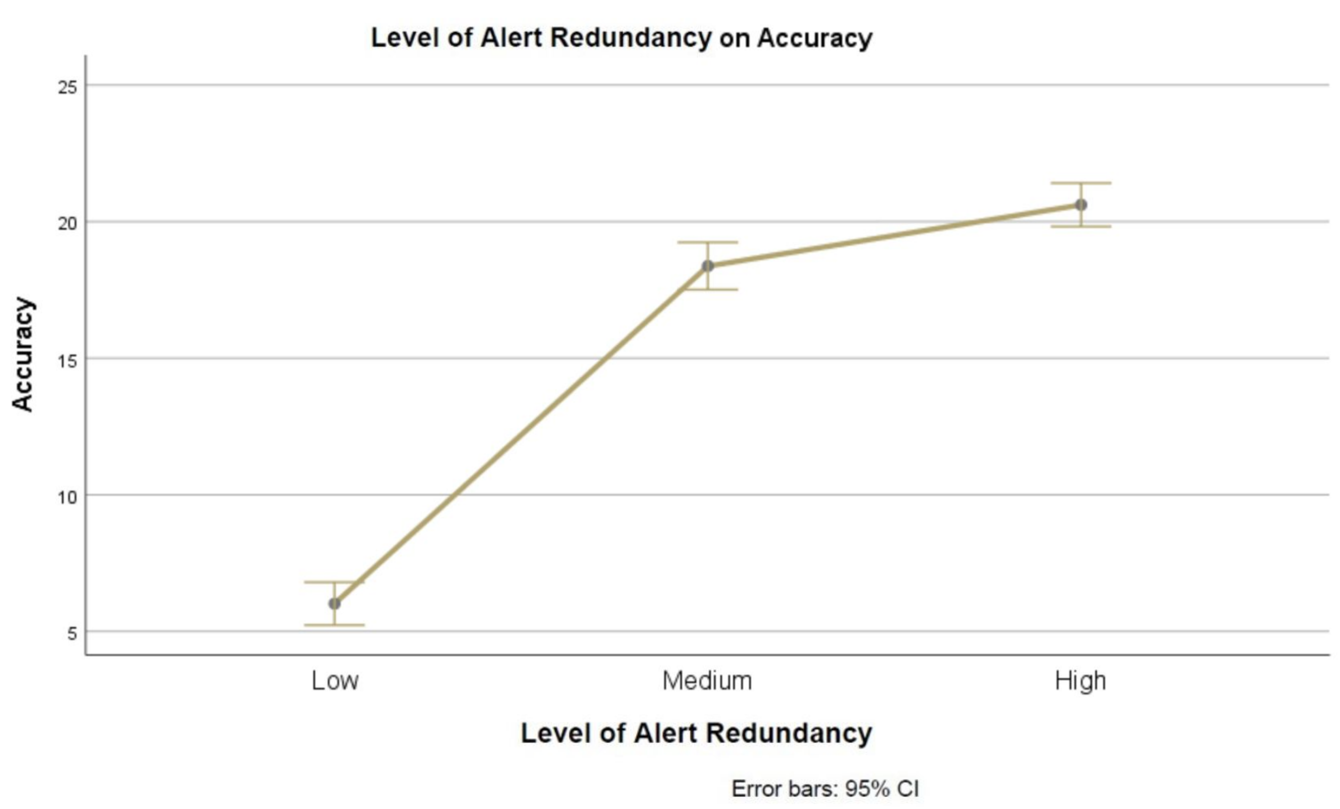
Informed Consent → Demographics → Task Instructions + Practice → Alphabetization + Alerts → Self-Report Measures → Debriefing

Results

Low Redundancy Leads to Decreased Accuracy of Alert Dismissals

There was a significant main effect off the level of alert redundancy on accuracy, $F(2, 198) = 381.88$, $p < .001$, $\eta^2 = .79$

Participants struggled to catalog the body of alerts when redundancy was low, as alerts repeated infrequently throughout the task in this condition.



Low Redundancy Leads to Increased Dwell Time Prior to Acting Upon Alerts

There was a significant main effect of alert redundancy on the average dwell time (time spent before reacting to an alert) , $F(2, 191) = 3.829$, $p = .02$, $\eta^2 = 0.04$.

When a more diverse array of alerts was shown to participants, they spent more time dwelling upon whether they had previously encountered a given alert.

*Four outliers >3SD above mean were removed from this analysis

High Redundancy Causes High Workload in the Variable Condition

There was a significant interaction between alert ratio and the level of alert redundancy on workload, $F(2, 189) = 3.21$, $p = .04$, $\eta^2 = 0.03$.

Participants had to remember which alerts were repeated and which alerts were new more often in the High Redundancy condition. This extra effort, along with an unpredictable (variable) alert ratio, may cause higher workload.

Future Directions + Implications

- **Future Direction:** Investigate in a **healthcare provider population** using alerts they experience in their work environment, as the generalizability of these results to providers may be limited.
- **Strength: Effect sizes** were calculated through this experiment with factors known to impact alert fatigue experimentally controlled and manipulated. Level of redundancy was found to have a robust effect on accuracy.
- **Future Direction:** Behavioral measures of alert fatigue (accuracy and dwell time) were uncorrelated with the subjective measure of alert fatigue → measures of subjective alert fatigue should be collected more often in informatics research.

Georgia Tech College of Sciences
School of Psychology

References

Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., & Kaushal, R. (2017). Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Medical Informatics and Decision Making, 17*(1), 1-9.

Ashrafi, S., Najafi Mehri, S., & Nehrir, B. (2017). Designing an alarm fatigue assessment questionnaire: Evaluation of the validity and reliability of an instrument. *Journal of Critical Care Nursing, 10*(4). doi: 10.5812/ccn.11647.

Cho, O. M., Kim, H., Lee, Y. W., & Cho, I. (2016). Clinical alarms in intensive care units: Perceived obstacles of alarm management and alarm fatigue in nurses. *Healthcare Informatics Research, 22*(1), 46-53.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143-156.

Elias, P., Peterson, E., Wachter, B., Ward, C., Poon, E., & Navar, A. M. (2019). Evaluating the impact of interruptive alerts within a health system: use, response time, and cumulative time burden. *Applied Clinical Informatics, 10*(05), 909-917.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

McGreevey III, J. D., Mallozzi, C. P., Perkins, R. M., Shelov, E., & Schreiber, R. (2020). Reducing alert burden in electronic health records: state of the art recommendations from four health systems. *Applied Clinical Informatics, 11*(01), 001-012.