Proceedings of the IISE Annual Conference & Expo 2024
A. Brown Greer, C. Contardo, J.-M. Frayret, eds.

Relationship between Conventional Workload Surrogates and VACP Assessments in Emergency Medical Services

Abstract ID: 5332

Jenna Ercolani [1], Laila Cure [1], Paul Misasi [2]

[1] Industrial, Systems and Manufacturing Engineering, Wichita State University, 1845 Fairmount St, Wichita, KS 67220, USA

[2] Sedgwick County EMS, 1015 Stillwell, Wichita, KS 67213, USA

Abstract

Workload surrogates commonly used in Emergency Medical Services (EMS) to evaluate the efforts of crewmembers throughout their workday have not been validated. This study investigated the relationship between workload assessments and surrogate metrics that can be calculated using conventional EMS dispatch data systems. Workload was assessed at random points in time during random shifts using the Visual, Auditory, Cognitive, Psychomotor (VACP) approach. Direct observation was used to assess the VACP scores of individual tasks commonly performed by ambulance crewmembers. Dispatch data was mapped to VACP score profiles through a trace-based simulation approach that adds random samples of tasks sequences and durations to the timestamps available in the EMS data. Pearson correlation and linear regression were used to quantify the relationship between the average time-weighted VACP workload and EMS metrics. Overall utilization and call response utilization explained time-weighted VACP the best, with strong positive correlations (0.88 and 0.89, respectively) and significant linear regression models (with R^2=0.79 and R^2=0.77, respectively). Call volume yielded the lowest correlation (0.61) and R^2 values (0.41) of the metrics studied, challenging its validity in fairly representing workload.

Keywords

VACP, NASA TLX, Emergency Medical Services, workload, paramedics

1. Introduction

This paper investigated the relationship of commonly used workload surrogates in emergency medical services (EMS) with assessments based on the Visual, Auditory, Cognitive, Physical (VACP) scale. While VACP assessments require time-consuming task analyses and simulation modeling, they can be interpreted as representing workload from a design perspective. The goal was to identify which easily quantifiable surrogate metric best represents the design workload of ambulance crewmembers.

Ambulance crewmembers, e.g., paramedics and emergency medical technicians (EMT) respond to 911 calls potentially involving medical treatment and transport. Typical shifts for these crews last 12 or 24 hours, depending on the system design and crews are often formed by two members per one ambulance [1]. While the frequency and nature of emergencies (i.e., call types) are random and highly variable, direct call response mainly involves driving to the call, treating the patient, transporting them to a hospital (if needed), and handing over the patient to the hospital staff. Other indirect work includes driving back to stations, call documentation, and ambulance cleanup. The timestamps of direct call response are often recorded in EMS dispatch systems, while indirect call work is often not known with certainty.

In EMS practice and research, workload is typically measured using call volume or unit hour utilization (UHU). However, these measures involve some questionable assumptions for this purpose. For example, the call volume assumes that all calls contribute equally to workload. UHU is a productivity measure established in 1983 [2] that is often misinterpreted as a utilization estimate (this interpretation implies that only transport calls affect workload and that responding to a single call takes exactly one hour). Given the increasing rates of burnout in EMS and the national paramedic shortage [3,4], it is important that these workload surrogates are studied in terms of their goodness of fit in representing the workload of ambulance crewmembers. This paper is organized as follows: First, we present a literature review of workload studies in EMS. Then, we introduce our methods, followed by the results and a discussion of the importance of our findings and opportunities for future research.

2. Related Research

The literature focusing on EMS workload has studied workload during specific tasks. Some studies have focused on perceived workload using the National Aeronautics and Space Administration (NASA) Task Load Index (NASA-TLX) [5-7], which is one of most widely used and validated measures of subjective workload [8]. Several studies have focused on physical workload using the Rating of Perceived Exertion and physical demand descriptions [9,10]. These studies found that different stretcher systems [9] and different types of call influenced the perceived physical demands [10]. Coffey et al. (2016) studied the perceived physical demands of EMS crewmembers over an entire shift. After measuring the medic's perceived ratings of clinical, physical, and emotional demand, this study found that these determinants of workload varied from call to call, and that a patient's acuity level was a determinant for a medic's exposure to high demands [11], potentially contradicting the use of call volume as a workload surrogate. No studies were found investigating EMS workload from a design perspective, i.e., representing an objective measure of workload consistent with the work that is expected from qualified crewmembers responding to assigned calls and performing all ensuing direct and indirect tasks.

The literature focusing on operational decision-making models in EMS, and that mentioned workload, involved developing optimization models for deployment [12-14], station location [15], hospital selection [16], and redeployment for coverage in a service area [17]. When workload was mentioned in these types of models, it usually referred to the overall call volume of the crew, or a utilization estimate based on direct call response time [18].

While there are many ways to measure workload [19], there are no universally validated workload metrics in general or for EMS. The visual, auditory, cognitive, psychomotor (VACP) method allows for workload assessments based on sequences of tasks and their observable characteristics as opposed to the subjective perceptions of the many individuals involved [20]. The VACP methodology builds upon the Multiple Resource Theory which was introduced by Wickens [21] and explained how mental workload is shared between the four dimensions corresponding to its name. The VACP method was developed in 1984 to help military organizations examine excessive workload in flight control missions [22]. Bierbaum et al. [23] updated the descriptor phrases used in the VACP scales. Since then, the model has been used to study missions in autonomous driving [24] and nuclear power plants [25]. VACP has also been used as an input to workload modeling software such as IMPRINT [19]. IMPRINT has been used in military operator design [26,27] and to study workload during multitasking driving scenarios [28]. While there is subjectivity in VACP assessments, these are assigned consistently by system designers or experts based on observations and through consensus. The assessments can be used to analyze observed or hypothetical sequences of tasks over time using simulation techniques, making it a useful measure in predictive modeling [19]. Furthermore, the VACP method has been shown to have some correlation with NASA TLX values of the same tasks [29-31] or has been verified by comparison of range estimates of the two values [24]. For these reasons, we selected the VACP method as the ground truth for design workload in the collaborating EMS system.

3. Methodology

To form our research dataset, we mapped routinely collected 911 dispatch data to possible task sequences (along with their durations and VACP scores) experienced by EMS crews within a study period. Then, we estimated the average time-weighted VACP score of the period based on those sequences, as well as the corresponding values of various EMS metrics.

3.1 Study Setting and Participants

The collaborating EMS system is a municipal ambulance service operating separately from the local fire departments and the 911 operator. In this system, 911 calls are mapped to one of nine response "priorities," which assign resources depending on the situations, with decreasing urgency as the priority number increases. Calls that present potentially life-threatening emergencies are given a priority 1 (P1). Each ambulance in the collaborating EMS service is typically operated by two paramedics who are assigned to one fixed geographical location (i.e., station) for their shift. At least one of the members is always a paramedic, while the other may be a paramedic or an EMT, depending on staffing constraints.

3.2 Workload Assessments

Researchers observed the tasks performed by EMS crews who had consented to the study at random dates, times, and stations during the study period. To use the VACP method, scores were recorded by observers using the verbal anchors in the VACP scales. Each workload dimension was scored on a scale of 0-7. The scores were summed to find the VACP score for each task [19]. This included tasks commonly recorded in dispatch data, as well as indirect work tasks (shift start, driving to station, and documentation activities). Observers then reached a consensus on standardized workflows and the corresponding VACP scores for the different tasks observed. The standard workflows considered variations in workload levels due to characteristics of calls. For example, the team agreed that a call assigned by dispatch as a priority 1 or 2, which involves lights and sirens when driving to the scene, resulted in a higher VACP score than calls with priorities 3 through 8, as these types of calls did not use lights and sirens. Researchers also identified the best fitting distributions of the durations of indirect call response tasks (i.e., those not recorded in the dispatch data).

NASA TLX surveys were given at the end of the observation period to each medic. This score was used to explore the relationship between the time-weighted VACP and NASA TLX values recorded at the same time of observation.

3.3 Research dataset formulation

We received dispatch data from the collaborating EMS system. This data set included records of every call to which ambulance crews were dispatched in the study period. Each record included call characteristics such as dispatch data (date, dispatch time, vehicle ID, medic identification, location details, and priority) and chart data (on-scene, transport, and in-service times, initial acuity of the call, the crewmember who performed documentation, transport destination).

We created a research dataset in which each record described each participating crewmember's workday up to the observation time. The attributes of the dataset included the date, time, station, subject identification, shift start time, total shift time, shift number (indicating day or night), call volume, the average time-weighted VACP estimate at the time, and one variable for each EMS metrics considered. The EMS metrics included the call volume (CV), defined as the count of calls to which the crew was dispatched in the period; the UHU, defined as the count of calls requiring transport divided by the period duration; the priority-stratified call volume (PS CV), defined as the count of calls of each response priority during the period; the initial-acuity-stratified call volume (IA CV), defined as the count of calls with different levels of initial acuity in the period (six initial acuities were considered: green, yellow, red, blue, black, and orange); call response utilization (CRU), defined as the sum of call durations recorded in dispatch data divided by the period length; overall utilization (OU), defined as the total estimated busy time (including direct and indirect call work) divided by the period duration; mean time between calls (MTBC), defined as the average time between end of one call and dispatch to next in the period; and the coefficient of variation of time between calls (CVTBC), defined as the standard deviation of time between calls in period divided by MTBC.

The average time-weighted VACP was estimated using a trace-based simulation coded in Python. The simulation modeled an individual crew member's entire workday based on observed dispatch data task patterns and Montecarlo samples of the corresponding indirect tasks from the assessed duration distributions. Recorded and simulated tasks were also assigned the task-level VACP values assessed during the observational study. The simulation was run for 30 iterations for each crewmember's workday up to the observation time and the average over the 30 replications of the time-weighted VACP value at that time was used as the "true" measure of workload.

3.4 Analysis

Pearson correlation was used to investigate the relationship between NASA TLX scores and VACP scores. Linear regression models were developed using each workload surrogate as a predictor and the VACP score as the response variable. The following control variables were used in the regression models: shift, licensure of crewmember (i.e., if the medic is a paramedic or EMT), partner's licensure, and lead (i.e., if the respondent is the lead medic on the shift).

4. Results

Data was collected from July 2022 to March 2023, and 73 records were obtained from 18 different ambulance stations. Average time-weighted VACP scores at the observation times ranged from 3.4 to 17.8, with no apparent relationship with the relative shift time. The Pearson correlation coefficient between NASA TLX scores and the VACP time-weighted average was 0.54, indicating a moderate positive correlation between the two measures of perceived workload ($p<0.001$).

Figure 1 shows the relationships between workload metrics and the average time-weighted VACP results. The OU and CRU had a high degree of positive correlation (0.89 and 0.88) with time-weighted VACP scores. Call volume (0.65), UHU (0.81), and CVTBC (0.69) also had positive linear correlations. MTBC had a moderate degree of negative correlation, at -0.68. All correlations were significant at the 1% level.

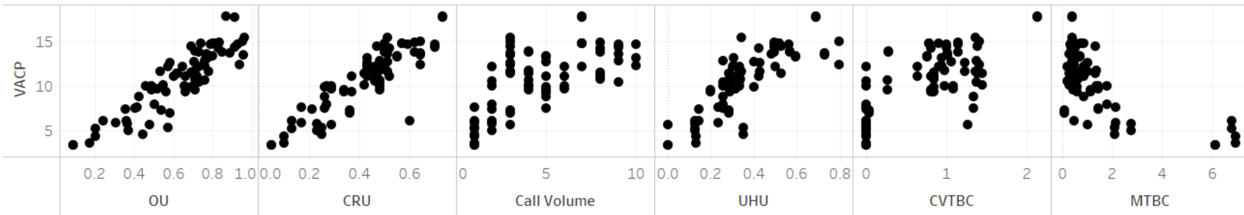

Figure 1. Pearson correlation analysis for workload metrics versus VACP.

The linear regression model with call volume as the predictor was significant at the 1% level but had the lowest R^2 (0.41), meaning that this model explains less than half of the variability in workload, as seen in Table 1. The linear models using the other surrogate workload measures as predictors were also all significant at the 1% level. The metrics that explained the highest amount of variability were OU and CRU, with R^2 values of 0.79 and 0.77, respectively. A model including all workload metrics as predictors had an R^2 of 0.83; however, the only significant predictors at the 10% level were UHU, OU, and CVTBC. UHU had an R^2 value of 0.69, followed by priority stratified call volume (R^2 = 0.49) and initial acuity call volume (R^2 = 0.46). The call volume model resulted in the poorest fit, followed by MTBC and CVTBC, both with an R^2 of 0.44. The control variables of shift, license, partner's license, and lead were not significant at the 10% level across all the models.

Table 1. Summary of linear regression outputs for workload metrics.

Model	**Adjusted R^2**	**Estimates**		**Coefficient P-Value**	
OU	0.79		14.71		2e-16
CRU	0.77		18.59		2e-16
UHU	0.69		16.93		2e-16
Priority-Stratified Call Volume *	0.49	P1 CV	1.00	P1 CV	6e-07
		P5 CV	0.79	P5 CV	0.04
		P8 CV	2.35	P8 CV	1e-06
IA CV *	0.46	Green	1.09	Green	8e-05
		Yellow	1.43	Yellow	7e-07
		Red	1.32	Red	0.02
MTBC	0.44		-1.43		5e-11
CVTBC	0.44		4.33		4e-11
Call Volume	0.41		3.85		0.03

**Only model predictors significant at the 5% level are shown for conciseness.*

5. Discussion

The correlation between the NASA TLX and time-weighted VACP scores at the same observation point suggests that VACP assessments based on historical dispatch data and the results of our task analyses are a reasonable measure of workload. Furthermore, the fact that control variables were not significant in the regression models confirms the consistency of the VACP as an objective measure of workload that depends only on the characteristics and durations of tasks regardless of the context or circumstances, such as shift or time of day.

Call volume had the lowest correlation with VACP and was the worst predictor among the candidate metrics studied. This finding challenges the convention of this metric as the best surrogate for workload. OU and CRU yielded the best results in the Pearson correlation and linear regression analyses. These results also suggest that the overall utilization (which incorporates the time spent in indirect work not commonly reported in dispatch systems) did not significantly add explanatory power to our models. The Pearson correlation was 0.01 higher for OU and only had an R^2 that was 0.02 higher than the CRU. These findings suggest that the call response utilization measure (which can be directly extracted from commonly available dispatch data systems) is almost as good at representing workload as the more realistic overall utilization measure which requires time-consuming task analysis and simulation development efforts.

The analysis showed that as MTBC increased, the workload assessments decreased. This indicates that when calls are spread out throughout the shift, the end-of-shift workload is lowered.

The limitations to this study lie in the development of the trace-based simulation and the data provided from the collaborating EMS system. As previously stated, calls can vary greatly in terms of required interventions, criticality, and mental demand. The way the information is captured in the dispatch data system and the way in which these variables were linked to workload may have influenced the results. No information on the medical interventions performed on patients during each call were included in this analysis. This means that every dispatched P1 call was assumed to have the same workload, which may not be true. In a similar way, a P8 transfer call can have a heavy physical workload not captured in the data used for this analysis. This could be a reason why the three different call volume metrics did not yield better results compared to the utilization estimates. The utilization estimates more accurately represent the medic's actual workday and work demand.

Future research may focus on identifying other (perhaps non-linear) functions of conventional EMS metrics to assess workload. An accurate workload assessment function could be incorporated into established optimization models to examine the tradeoffs between operational and workload outcomes.

6. Conclusion

The conventional metrics of call volume and UHU in EMS have been challenged as workload surrogates by investigating their relationship with VACP assessments. According to the VACP assessments, our study suggests that a utilization estimate based on direct call response tasks commonly recorded in emergency dispatch data systems is the best indicator of workload among the measures studied. The conventional measure of call volume had the lowest correlation to workload and the lowest R^2 in its linear model. Besides suggesting alternative VACP workload surrogates, this finding suggests that further research is needed to identify a valid measure representing the experience of first responders in EMS systems. Then, the validated workload surrogates could be used to develop strategies to alleviate or balance workload in EMS operations.

Acknowledgements

We would like to thank Sedgwick County EMS for their assistance and collaboration in providing the data for this research. This work was supported by the National Science Foundation, award number 2138995.

References

[1] Brown, L. H., Owens, C. F., March, J. A., & Archino, E. A., "Does Ambulance Crew Size Affect On-Scene Time or Number of Prehospital Interventions?," *Prehospital and Disaster Medicine,* vol. 11, no. 3, pp. 214–217, 1996, doi:10.1017/S1049023X00042977.

[2] Stout, J. L., "System status management: The strategy of Ambulance Placement," *JEMS*, vol. 8, no. 5, pp. 22–23, 1983.

[3] K. Snyder, "Ambulance services face national paramedic shortage," The Blade, 2019, [Online]. Available: https://www.ems1.com/paramedic-jobs-and-careers/articles/ambulance-services-face-national-paramedic-shortage-QAINB7yNEngfu5ID/. [Accessed January 31, 2024].

[4] R. P. Crowe, J. K. B. R. E. C. A. R. P. S. A. R. and S. E. O.-M., "Association of Burnout with Workforce-Reducing Factors among EMS Professionals," *Prehospital Emerg. Care*, vol 22, no. 2, pp. 229–236, 2018

[5] Keunecke, J. G., Gall, C., Birkholz, T., Moritz, A., Eiche, C., & Prottengeier, J., "Workload and influencing factors in non-emergency medical transfers: A multiple linear regression analysis of a cross-sectional questionnaire study," *BMC Health Services Research*, vol. 19, no. 1, 2019, doi:10.1186/s12913-019-4638-4.

[6] Prottengeier, J., Keunecke, J. G., Gall, C., Eiche, C., Moritz, A., & Birkholz, T., "Single mission workload and influencing factors in German prehospital emergency medicine - A nationwide prospective survey of 1361emergency missions," Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, vol. 27, no. 1 ,2019, doi:10.1186/s13049-019-0650-2.

[7] Schneider, F., Martin, J., Schneider, G., & Schulz, C. M., "The impact of the patient's initial NACA score on subjective and physiological indicators of workload during pre-hospital emergency care," PLoS ONE, vol. 13, no. 8, 2018, doi:10.1371/journal.pone.0202215.

[8] NASA, "NASA TLX: Task Load Index," 2020. [Online]. Available: https://humansystems.arc.nasa.gov/groups/TLX/. [Accessed January 20, 2024].

[9] Hulldin, M., Kängström, J., Andersson Hagiwara, M., & Claesson, A., "Perceived exertion using two different EMS stretcher systems, report from a Swedish study," *The American Journal of Emergency Medicine*, vol. 36, no. 6, pp. 1040–1044, 2018, doi:10.1016/j.ajem.2018.02.023.

[10] Morales, L., McEachern, B. M., MacPhee, R. S., & Fischer, S. L., “Patient acuity as a determinant of paramedics’ frequency of being exposed to physically demanding work activities.” Applied Ergonomics, vol. 56, pp. 187–193, 2016, doi:10.1016/j.apergo.2016.04.003.

[11] Coffey, B., MacPhee, R., Socha, D., & Fischer, S. L., “A physical demands description of paramedic work in Canada,” *International Journal of Industrial Ergonomics,* vol. 53, pp. 355–362, 2016, doi:10.1016/j.ergon.2016.04.005.

[12] Bandara, D., Mayorga, M. E., & Mclay, L. A., “Priority dispatching strategies for EMS systems,” *Journal of the Operational Research Society*, vol. 65, no. 4, pp. 572–587, 2014, doi:10.1057/jors.2013.95.

[13] Enayati, S., Özaltın, O. Y., Mayorga, M. E., & Saydam, C., “Ambulance redeployment and dispatching under uncertainty with personnel workload limitations,” *IISE Transactions*, vol. 50, no. 9, pp. 777–788, 2018, doi:10.1080/24725854.2018.1446105.

[14] Kergosien, Y., Bélanger, V., Soriano, P., Gendreau, M., & Ruiz, A., “A generic and flexible simulation-based analysis tool for EMS management,” *International Journal of Production Research*, vol. 53, no. 24, pp. 7299–7316, 2015, doi:10.1080/00207543.2015.1037405.

[15] Li, X., Zhao, Z., Zhu, X., & Wyatt, T., “Covering models and optimization techniques for emergency response facility location and planning: A review,” *Mathematical Methods of Operations Research*, vol. 74, no. 3, pp. 281–310, 2011, doi:10.1007/s00186-011-0363-4.

[16] Caicedo Rolón, A. J., & Rivera Cadavid, L., “Hospital selection in emergency medical service systems: A literature review,” *Gerencia y Políticas de Salud*, vol 20, 2021, doi:10.11144/javeriana.rgps20.hsem.

[17] Enayati, S., Mayorga, M. E., Toro-Díaz, H., & Albert, L. A., “Identifying trade-offs in equity and efficiency for simultaneously optimizing location and multipriority dispatch of ambulances,” *International Transactions in Operational Research*, vol. 26, no. 2, pp. 415–438, 2019, doi:10.1111/itor.12590.

[18] Aboueljinane, L., Sahin, E., & Jemai, Z, “A review on simulation models applied to emergency medical service operations,” *Computers and Industrial Engineering*, vol. 66, no. 4, pp. 734–750, 2013, doi:10.1016/j.cie.2013.09.017.

[19] Rusnock, C. F., & Borghetti, B. J., “Workload profiles: A continuous measure of mental workload,” *International Journal of Industrial Ergonomics*, vol. 63, pp. 49–64, 2018, doi:10.1016/j.ergon.2016.09.003.

[20] Mitchell, D. K., “Mental Workload and ARL Workload Modeling Tools 20000518 033 ÜTIC QUALITY INSPECTED Ö,” *Army Research Library,* 2000.

[21] Wickens, C. D., “Multiple resources and performance prediction,” *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159–177, 2002, doi:10.1080/14639220210123806.

[22] McCracken, J. H., & Aldrich, T. B., “Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals,” 1984.

[23] Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B., “Task Analysis of the UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model,” *US Army Research Institute,* vol. 1., 1989.

[24] Chen, W., Sawaragi, T., & Horiguchi, Y., “Measurement of Driver’s Mental Workload in Partial Autonomous Driving,” *IFAC-PapersOnLine*, vol. 52, no. 19, pp. 347–352, 2019, doi:10.1016/j.ifacol.2019.12.083.

[25] Zhang, Y., Wen, H., Dai, X., Liang, J., Xu, Z., Xue, K., & Yang, M., “Research on modeling in operator mental workload based on VACP method,” *Quality and Reliability Engineering International*, 2022, doi:10.1002/qre.3189.

[26] Hunn, B. P., Schweitzer, K. M., Cahir, J. A., & Finch, M. M., “IMPRINT Analysis of an Unmanned Air System Geospatial Information Process,” 2008.

[27] Mao, M., Xie, F., Hu, J. jun, & Su, B., “Analysis of workload of tank crew under the conditions of informatization,” *Defence Technology*, vol. 10, no. 1, pp. 17–21, 2014, doi:10.1016/j.dt.2013.12.008.

[28] Kandemir, C., Handley, H. A. H., & Thompson, D., “A workload model to evaluate distracters and driver’s aids,” *International Journal of Industrial Ergonomics*, vol: 63, pp. 18–36, 2018, doi:10.1016/j.ergon.2016.09.004.

[29] Borghetti, B. J., Giametta, J. J., & Rusnock, C. F., “Assessing Continuous Operator Workload with a Hybrid Scaffolded Neuroergonomic Modeling Approach,” *Human Factors*, vol. 59, no. 1, pp. 134–146, 2017, doi:10.1177/0018720816672308.

[30] Rusnock, C., Borghetti, B., & McQuaid, I., *“Objective-Analytical Measures of Workload – the Third Pillar of Workload Triangulation?,”* in *Proceedings of 2015 International Conference on Augmented Cognition,* pp. 124–135, 2015, doi:10.1007/978-3-319-20816-9_13.

[31] Wang, P., *“A colored petri nets based workload evaluation model and its validation through Multi-Attribute Task Battery-II,” Applied Ergonomics,* vol. 60, pp. 260-274, 2017, doi:10.1016/j.apergo.2016.11.013.