Test-Retest of the 5-Minute Psychomotor Vigilance Test

Katharine Dennison
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Medicine and Health Sciences Commons

Recommended Citation
Dennison, Katharine, "Test-Retest of the 5-Minute Psychomotor Vigilance Test" (2021). All Graduate Plan B and other Reports. 1560. https://digitalcommons.usu.edu/gradreports/1560

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
TEST-RETEST OF THE 5-MINUTE PSYCHOMOTOR VIGILANCE TEST

by

Katharine Dennison

A plan B research project submitted in partial fulfillment of the requirements for the degree of

Master’s of Science

In

Health and Human Movement

Approved:

____________________________  ______________________________
Brennan Thompson, Ph.D.        Talin Louder, Ph.D.
Major Professor                Committee Member

____________________________
Ray Joslyn, Ph.D.
Committee Member

UTAH STATE UNIVERSITY

Logan, Utah

2021
INTRODUCTION

It is common in occupational domains, such as in the healthcare field, that time-dependent reactions must be made to avoid potentially damaging outcomes. For instance, the ability to react quickly can increase the likelihood of avoiding a fall from a slip or trip or avoiding being struck by a moving object. Also, working in a healthcare environment involves decisions regarding patients’ treatments that must be made quickly and accurately under rapidly changing circumstances. For instance, the correct dosage of medication needs to be administered in a prescribed time frame or emergency surgery decisions need to be made to save a patient’s life. These quick decisions and/or reactions are often the difference between sustaining an injury or health hazard versus avoiding them, and can even make the difference between life and death.

Nurses and doctors who work long twelve-hour shifts are often performance-impaired by fatigue towards the end of their shifts (Thompson, 2021; Thompson, 2019; Behrens et al, 2019). For instance, a study by Thompson (2019) found fatigue-related performance impairments in nurses after just one 12-hour shift, with compounding effects of fatigue observed after multiple, successive work shifts. This data is further validated by the findings of Geiger-Brown et al. (2021) who reported that nurses and healthcare workers often struggle with fatigue and poor sleep quality after a 12-hour shift (Geiger-Brown et al., 2021). Thus, fatigue-related impairments are common in occupational settings, especially those where demanding work schedules are prevalent (e.g., long shifts, compressed schedules, physical or mentally taxing work), with the consequences of such impairments including a variety of negative outcomes (including physical, mental, and economical) that may impact the healthcare worker and patient, as well as the organization and society at large (Thompson 2021).
The Psychomotor Vigilance Test (PVT) is widely used as an assessment of one’s neurocognitive capacities pertaining to sustained alertness. The PVT returns measurements of reaction time, false starts, and performance lapses over a several minute period from tasks involving repeated reactions to frequent stimuli (Ferris et al., 2021; Behrens et al., 2019). The test has been demonstrated to be a valid and useful tool for identifying fatigue- and/or sleep-related impairments in a number of various populations (Thompson 2019; Thompson 2021; Wilson et al. 2019; Behrens et al. 2019; Basner and Dinges, 2012; Van Dongen et al., 2003). A three-minute PVT was used by Behrens et al (2019) which observed delayed and longer reaction times with more lapses in female night shift workers as compared to the day workers. Other recent studies have also demonstrated this elongated reaction time effect as assessed via the PVT in female only populations as a result of working rigorous work shifts in hospital-based settings (Thompson, 2019; Thompson et al. 2016; Wilson et al. 2018). Thus, the PVT is widely used and accepted as a valid test for monitoring fatigue- and sleep-related impairments.

The PVT was originally developed as a ten-minute test, and this is the duration for which it has mostly been used to assess fatigue- and sleep-related impairments over approximately the last three decades. While the 10-minute PVT has been demonstrated to be reliable (Sunwoo et al., 2012), recently, a number of studies have used the PVT at shorter durations, such as with 5-minute (Thompson et al., 2016; Thompson et al., 2019; Wilson et al, 2010) and even 3-minute (Behrens et al., 2019; Matsangas et al., 2019; Basner et al., 2011) testing times. The reason for conducting the PVT at shorter than traditional (< 10 minutes) durations is the practical utility of doing so (Basner and Dinges, 2012; Basner, Mollicone, and Dinges, 2011). It is rather burdensome, and thus impractical, to repeatedly conduct a 10-minute, non-engaging (e.g., rather mundane) PVT task on a relatively large population of workers or subjects, and so
shorter versions have been implemented to increase the practicality of the test. The rationale for the use of shorter duration PVT finds support from the findings of Basner, Mollicone, and Dinges (2011) who found that the sensitivity and validity of a 3-minute version of the PVT were comparable with the more traditional 10-minute test in a multi-day sleep deprivation study. Another study by this group (Basner and Dinges, 2012) substantiated the use of a truncated PVT by demonstrating that a 95.7% accuracy was observed for correctly classifying the PVT data with an average PVT test duration of 6-minutes. These studies confirm that the PVT can be highly sensitive and valid at reduced durations of 40 – 70% that of the traditional 10-minute assessment duration. Other studies (Roach, Dawson, and Lamond, 2006; Loh et al., 2004) have also successfully validated the use of shorter PVT durations, with Loh et al. (2004) concluding that it may be a “viable alternative to the 10-minute PVT.”

These shorter versions of the PVT have been also been demonstrated to be quite sensitive to fatigue-based changes in PVT performance. Recent studies conducted by our research team (Thompson et al., 2016; Thompson 2019) demonstrated that the five-minute PVT duration is valid for identifying fatigue-induced reaction time and error-based increases in female nurses where fatigue was induced via demanding work schedules. Thus, the five-minute PVT has been demonstrated to be proficient and successful in field-based environments for capturing fatigue impairments in working populations in the context of a repeated measures testing scheme.

The ideal duration of a shortened PVT has yet to be unequivocally established. However, the use of the 5-minute duration version of the test over other durations may find support by the findings of Loh et al. (2004) and Roach, Dawson, and Lamond (2006). For example, in a sleep restricted study, Loh et al. (2004) compared changes in PVT performance during a standard 10-minute PVT with the first 5-minutes and 2-minutes of the PVT. They found that while the first 5-
minutes and 2-minutes were both sensitive to the effects of sleep loss, the magnitude of changes for the first 2-minutes was substantially less than that of the first 5-minutes. They concluded that a 5-minute duration provides a more accurate measure of the changes versus a 2-minute duration. Findings by Roach, Dawson, and Lamond (2006) were similar, such that the 5-minute PVT condition showed a higher correlation to the 10-minute PVT than the 90-second PVT. Both of these authors suggested that the 5-minute PVT was the better substitute for the 10-minute PVT compared to the shorter (< 5-minute) versions. This indicates that there is supportive evidence to shorten the PVT duration and still have a high level of sensitivity and/or accuracy.

Despite being an increasingly prevalent and useful test, the test-retest evaluation using traditional reliability statistics of the 5-minute PVT version has not been examined nor reported in an adult female working-type population (e.g., other than solely college-age and/or students). In fact, we are aware of only a single study that has reported reliability of the 5-minute PVT. A study by Wilson et al. (2010) examined the reliability with a 5-minute PVT duration in male primary school-aged youth (age = 11.8 years). The authors reported a moderately high reliability with intraclass correlation coefficients (ICCs) between .59 (lapses) and .84 (mean reaction time; RT) (Wilson et al., 2010). Thus, the reliability of the 5-minute PVT would be important to investigate and quantify in order to inform researchers and clinicians who wish to use this shorter, more practical version of the PVT. Providing a comprehensive evaluation of the 5-minute PVT reliability would provide an important framework for understanding the consistency of the test, and help give context to its scientific quality and utility to help provide support for its continued use. Therefore, the purpose of the present study was to investigate the test-retest reliability of the 5-minute PVT for a multitude of reaction time- and error-based outcome
measures in a population of working age females. A secondary aim was to report the reliability of a sleepiness scale that is commonly administered with the PVT test.

METHODS

Participants

Thirteen working age (20 – 63 years) females (mean ± SD: age = 39.1 ± 14.6 years, height = 167.2 ± 5.1 cm, weight = 66.6 ± 6.7 kg) volunteered to participate in this investigation. Participants were eligible for the study if they were between 18 and 65 years of age, had no lower body musculoskeletal injuries within the last one-year period, and did not have any neuromuscular disorders (e.g., Parkinson’s). Participants were instructed to make no changes to their lifestyle during the course of the study and were instructed to not participate in any rigorous physical activities or exercise within the 24-hour period prior to the study and between sessions. Additionally, participants were instructed to keep their caffeine intake consistent on the days of the study trials, so as to not disrupt their ordinary routine, which could have an unnatural effect on the study outcomes. This study was approved by the university’s institutional review board, and all participants read and signed an informed consent form prior to any data collection.

Experimental Design

This study employed a repeated measures design to evaluate the reliability of the 5-minute PVT. Participants reported to the lab on two occasions, separated by 24 – 72 hours. The first session involved questioning to confirm eligibility for the study, provision of written informed consent, measurement of height and weight, and a run through (practice trial) of a truncated version of the PVT (1-minute) to instruct and familiarize the participants on the test. A
A 5-minute break was provided after the familiarization period before completing the 5-minute PVT. The second visit only involved the 5-minute PVT and the self-recorded sleepiness scale.

**PVT Administration**

The PVT procedures were implemented in accordance with a protocol described previously (Thompson 2019). The PVT used a freely available PC-PVT software platform as validated by Khitrov et al. (2014). The testing was conducted with a PC (HP Z220, HP Inc., Palo Alto, CA) and an optical gaming mouse (Razer Abyssus, Razer, Carlsbad, CA). The duration of the PVT was five minutes and participants were asked to turn off all personal electronic devices and were prohibited from having any other persons in the room during testing. The PVT variables were calculated from the software and included the mean (MRT), fastest and slowest 10% of reaction times (F10RT% and S10RT%), standard deviation of reaction time (SDRT), as well as the number of errors (ERR; which included combined major and minor lapses, anticipations, and false starts). In addition, the generic sleepiness scale was also assessed, which asked the participants “how sleepy are you” on a scale of 1 – 10 immediately prior to beginning the PVT.

**Statistical Analysis**

Descriptive statistics were reported as means and SD for all variables. Paired samples t-tests were performed to examine the systematic variability for all variables. For relative consistency evaluation, the intraclass correlation coefficient (ICC) was used. Absolute consistency was determined from the standard error of measurement (SEM) and minimal difference (MD) to be considered real statistics.

The ICC model “2,1” was used as described by Shrout and Fleiss (1979) and was selected on account of its generalizability to other laboratories and testers (Weir 2005). The ICC, SEM,
and MD statistics were calculated from equations provided by Shrout and Fleiss (1979) and Weir (2005). The ICCs were evaluated based on a reliability scale where ICCs < 0.50 displays poor reliability, ICCs between 0.50 – 0.77 show moderate reliability, and ICCs > 0.75 show high reliability (Koo & Li, 2015). An alpha level of P < 0.05 was used to determine statistically significant outcomes.

Results

The descriptive statistics (mean ± SD) are presented in Table 1. The reliability statistics are presented in Table 2. Reliability statistics revealed that there was either poor, or undeterminable (as described in the discussion section below) reliability in regards to minor lapses, major lapses, anticipations, SDRT, and S10RT% (ICCs < 0.50). There was moderate reliability in regards to false starts, ERR, and sleepiness scale (ICCs 0.50 – 70). Finally, there was high reliability in regards to MRT and F10RT% (ICC > 0.7). There was systematic error found for the MRT, SDRT, and the S10%RT (P = 0.001 – 0.01). There was no systematic error for any of the other variables (P > 0.05).
Table 1. Descriptive statistics showing the mean and the standard deviation (SD) of the variables in the 5-minute Psychomotor Vigilance Test for test trials (T) 1 and 2.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Trial</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor Lapses (# ≥ 500 ms)</td>
<td>T-1</td>
<td>0.231</td>
<td>0.439</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Major Lapses (# ≥ 1000 ms)</td>
<td>T-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Anticipations (# &lt; 100 ms)</td>
<td>T-1</td>
<td>0.077</td>
<td>0.277</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>False Starts (#)</td>
<td>T-1</td>
<td>1.462</td>
<td>4.409</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>1.154</td>
<td>2.193</td>
</tr>
<tr>
<td>ERR (#)</td>
<td>T-1</td>
<td>1.769</td>
<td>4.362</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>1.154</td>
<td>2.193</td>
</tr>
<tr>
<td>MRT (ms)</td>
<td>T-1</td>
<td>267.836</td>
<td>29.667</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>254.821</td>
<td>29.019</td>
</tr>
<tr>
<td>SDRT (ms)</td>
<td>T-1</td>
<td>53.425</td>
<td>10.410</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>42.342</td>
<td>11.061</td>
</tr>
<tr>
<td>F10RT% (ms)</td>
<td>T-1</td>
<td>204.350</td>
<td>20.092</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>203.434</td>
<td>22.365</td>
</tr>
<tr>
<td>S10RT% (ms)</td>
<td>T-1</td>
<td>380.469</td>
<td>41.923</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>343.085</td>
<td>39.494</td>
</tr>
<tr>
<td>Sleepiness</td>
<td>T-1</td>
<td>2.538</td>
<td>1.450</td>
</tr>
<tr>
<td></td>
<td>T-2</td>
<td>2.615</td>
<td>1.261</td>
</tr>
</tbody>
</table>

ERR = total errors including minor and major lapses, anticipations and false starts; MRT = mean reaction time; SDRT = standard deviation of reaction times over the 5-minute PVT trial; F10RT% = fastest 10% of reaction times; S10RT% = slowest 10% of reaction times; sleepiness = how tired the subjects were on a scale of 1 – 10, as assessed immediately prior to taking the PVT. T-1 = trial 1; T-2 = trial 2
Table 2. Reliability statistics for the Psychomotor Vigilance Test (PVT) for minor lapses, major lapses, anticipations, false starts, total errors, mean reaction time, standard deviation of reaction time, fastest 10%, slowest 10%, and the sleepiness scale.

<table>
<thead>
<tr>
<th>Variables</th>
<th>P-value</th>
<th>ICC 2,1</th>
<th>95% CI for ICC</th>
<th>SEM</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor Lapses</td>
<td>0.082</td>
<td>NA</td>
<td>N/A</td>
<td>0.310</td>
<td>0.860</td>
</tr>
<tr>
<td>Major Lapses</td>
<td>NA</td>
<td>NA</td>
<td>N/A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Anticipations</td>
<td>0.337</td>
<td>0.283</td>
<td>N/A</td>
<td>0.196</td>
<td>0.544</td>
</tr>
<tr>
<td>False Starts</td>
<td>0.728</td>
<td>0.615</td>
<td>0.11-0.87</td>
<td>2.206</td>
<td>6.114</td>
</tr>
<tr>
<td>ERR</td>
<td>0.495</td>
<td>0.593</td>
<td>0.09-0.86</td>
<td>2.227</td>
<td>6.174</td>
</tr>
<tr>
<td>MRT</td>
<td>0.010</td>
<td>0.794</td>
<td>0.26-0.94</td>
<td>10.825</td>
<td>30.005</td>
</tr>
<tr>
<td>SDRT</td>
<td>0.001</td>
<td>0.439</td>
<td>-0.11-0.8</td>
<td>6.253</td>
<td>17.332</td>
</tr>
<tr>
<td>F10%RT%</td>
<td>0.800</td>
<td>0.830</td>
<td>0.53-0.95</td>
<td>9.031</td>
<td>25.032</td>
</tr>
<tr>
<td>S10%RT%</td>
<td>0.003</td>
<td>0.431</td>
<td>-0.1-0.78</td>
<td>25.763</td>
<td>71.411</td>
</tr>
<tr>
<td>Sleepiness</td>
<td>0.819</td>
<td>0.636</td>
<td>0.14-0.87</td>
<td>0.840</td>
<td>2.328</td>
</tr>
</tbody>
</table>

ERR = total errors including minor and major lapses, anticipations and false starts; MRT = mean reaction time; SDRT = standard deviation of reaction times over the 5-minute PVT trial; F10%RT% = fastest 10% of reaction times; S10%RT% = slowest 10% of reaction times; sleepiness = how tired the subjects were on a scale of 1–10, as assessed immediately prior to taking the PVT. ICC 2,1 = intraclass correlation coefficient using model 2,1; 95% CI = confidence interval for the ICC data; SEM = standard error of measurement; MD = minimal difference needed to be considered real.

Discussion

This study focused on reporting reliability data for the 5-minute PVT. After analyzing the reliability data, MRT and the F10%RT were found to be the most reliable variables. The variables with medium reliability were false starts, ERR, and sleepiness scale. Finally, the variables with the least amount of reliability were minor and major lapses, anticipations, SDRT, and the S10%RT. This data that was collected is applicable not only to shift workers, but could potentially apply to anyone with sleep deprivation/fatigue that is comparable to that of shift workers i.e. college students, new parents, etc.

The variables that showed the highest reliability were MRT and the F10%RT with ICCs of 0.794 and 0.830, respectively. The classification of the reliability data herein is based on the
ICC reliability ranges as described in Koo & Li (2015). Our observation of moderately high reliability for MRT is in agreement with the finding of Wilson et al. (2010), who reported an ICC of 0.84. MRT may be considered one of the most commonly reported PVT variables. The consistency of MRT reliability is encouraging given that Wilson et al. (2010) included a very different sample population, as they analyzed data corresponding to male primary aged children in comparison with the female young adults included in the present study. Interestingly, F10RT% was modestly more reliable than MRT. It is notable that MRT showed systematic error (P = .010), with significant improvement observed from trial 1 to trial 2, despite the subjects undergoing a brief (1-minute) familiarization. Systematic error was not observed for F10RT%.

Two conclusions may be drawn from this rather novel finding. First, when investigations are primarily focused on examining and reporting MRT, more time may need to be devoted to familiarization in order to avoid systematic error resulting from a learning effect. Second, the F10RT% variable may be a more time efficient, as well as more reliable variable compared to MRT given that it did not exhibit systematic error. The present findings may suggest that the F10RT% variable is perhaps currently an under-utilized PVT outcome parameter and we suggest that this variable can be recommended with more test-retest related confidence when assessing PVT research measurements in the future. Moreover, when circumstances prevent considerable familiarization procedures, it may be recommended to focus PVT-based outcomes on the F10RT% variable rather than MRT. Further research is needed in other populations (e.g., elderly, working males etc.) to confirm whether these reliability results are also applicable across a broad range of populations.

False starts, ERR, and sleepiness were found to be moderately reliable. These variables were not reported in the Wilson et al. (2010) study, and there is a lack of comparative reliability
statistics in the literature. Sleepiness showed an ICC of 0.636 and no systematic error. Sleepiness is a rather novel variable that it is assessed prior to completing the PVT test on the software application used in the present study. Specifically, subjects were asked to report their level of sleepiness on a scale of 1-10 before they started the 5-minute PVT. The present finding implies that the general sleepiness levels reported by the subjects were moderately reliable and this may be a metric that shows a mirroring effect in the subjects’ performance on the 5-minute PVT. This variable may be useful in the future as a moderately reliable variable to use when assessing one’s sleepiness factor in conjunction with objective PVT-based variables. However, more research is warranted involving self-reported subjective scores on sleepiness and/or mood which are then compared to the PVT. False starts and ERR had ICCs of 0.615 and 0.593, respectively. There was no systematic error for both false starts and ERR. ERR, specifically, is a composite score of all types of potential errors, which included minor and major lapses, anticipations, and false starts. Thus, false starts and ERR are moderately reliable indicators of the predisposition of subjects to make errors on the 5-minute PVT which, in theory, increases with sleepiness and fatigue.

Minor and major lapses, anticipations, SDRT, and the S10%RT were found to have poor reliability. The SDRT represents the standard deviation of the reaction time responses over the 5-minute PVT period and the S10%RT represents the slowest 10% of all the responses and thus both are time domain related metrics (as opposed to error related). The reasons for the poor reliability on these two metrics are unknown, however, we speculate that the slower portion of the responses are reflective of perhaps rare events such as a lapse in attention to the task that may yield one or two outliers which would not necessarily be the type of thing that is replicated in subsequent tests. SDRT and the S10RT% also showed systematic error, further indicating the
problems associated with their use in a minimal familiarization design such as the one implemented in this study.

Interestingly, the present study did not report any ICCs for minor or major lapses because these events largely did not occur. For instance, there was only 1 minor lapse recorded in the first trial and no minor lapses in the second trial and no major lapses in either trial. This differs from the previously recorded ICC of .59 for lapses found by Wilson et al. (2010). The different populations used in the two studies or an unknown factor could possibly explain this difference. It is likely that these types of errors (lapses) become exponentially more likely to occur with even a modest amount of fatigue. Since the purpose of the study was reliability, there was no fatigue intervention nor any other condition that would have unfavorably impacted the variables. It appears that at a well rested, baseline status the lapses are not a factor in the PVT. Further research is needed to look more in depth at lapses in the 5-minute PVT and how the reliability statistics of lapses are impacted by varying conditions of the subjects (such as with different levels of fatigue or sleep restriction). Based on the results of this study, we do not recommend using the SDRT and S10RT% variables with the 5-minute PVT, and lapses may also not be recommended for this duration of test in this population, although we caution per the point above regarding the way in which the lapses variables may become more relevant, and better able to be assessed, when fatigue is present versus when it is not.

Although there are many novel aspects of our study which have been previously mentioned, there are areas and variables where more research is needed to gain more data and understanding in regards to the 5-minute PVT and its reliability to asses fatigue and sleepiness levels. The main limitation of this study was that only women were included in the testing process and so these data may not generalize to other populations. More research is needed to
study adult male shift workers as well as older adult populations especially those with clinical conditions (i.e., chronic fatigue syndrome). There is also the opportunity to do more research with the female adolescent population as Wilson et al. (2010) only studied male subjects. Another limitation of this study is that our population size was not very large. Thirteen individuals may be enough to get accurate data for the reliability of the 5-minute PVT, especially for some of the error-based metrics, and so future studies should include more subjects so that this version of the test can be further substantiated from a reliability perspective. It was rather unexpected that MRT improved between the first and second trials. There was apparently a learning curve between trials 1 and 2 for MRT, despite subjects performing a minimal PVT familiarization (1-minute). More work is needed to determine optimal familiarization procedures that balance both the practicality of test administration and variable reliability.

Overall, the main findings in this study suggest that MRT and the F10%RT variables are the most reliable outcome measures in the 5-minute PVT test in working adult females. The F10%RT variable may be an under-utilized metric as produced the highest reliability coupled with no systematic error for the 5-minute PVT in a minimal familiarization testing model. When assessing errors is of interest particularly in the context of the 5-minute PVT, the present study showed that the two most reliable error-related metrics were false starts and total errors (ERR) (consisting of a summation of all error-related events into one composite metric) which give a moderate level of reliability. Finally, under normal conditions (baseline, or non fatigued conditions) minor and major lapses are not recommended based on these data for the 5-minute PVT, but more research is needed to confirm the reliability of these variables when fatigue is present, which may yield more events. Finally, the S10RT% and SDRT variables are not recommended for the 5-minute PVT given their poor reliability.
References


http://dx.doi.org/10.5665/sleep.1620


Kay, M., Rector, K., Consolvo, S., Greenstein, B., Wobbrock, J.O., Watson, N. F., & Kientz,


Performance and sleepiness in nurses working 12-h day shifts or night shifts in a community hospital. *Accident Analysis and Prevention*,

http://dx.doi.org/10.1016/j.aap.2017.09.023