

6.0 Results of Feedback from FMT Technologies: Canada Phase

As described above, Study Phase 1 (data collection in 2002) took place under Canadian hours-of-service, and involved a Canadian trucking company (Challenger Motor Freight, Ontario, Canada) in which volunteer drivers operated single tractor-trailer units with sleeper berths, and approximately 26% of their driving was conducted during nighttime hours (74% in daylight hours). Study phase 2 (data collection in 2003) took place under U.S. hours-of-service, and involved a U.S. trucking company (Con-Way Central Express, Ann Arbor, Michigan, U.S.) in which volunteer drivers operated tandem tractor-trailer units without sleeper berths, and approximately 93% of their driving was conducted during nighttime hours (7% in daylight hours). The difference between Canadian and U.S. trucking companies were in part a function of which companies agreed to be part of the study, as well as our goal to expressly study companies in which night driving was both a minority (Study Phase 1) and a majority (Study Phase 2) of trucking operations. For these reasons the Canada study phase and U.S. study phase were analyzed separately for the effects of FMT FEEDBACK on driving and alertness outcomes, before being combined (later sections). This section presents the results for the Canada Study Phase 1. As described above, a total of $n = 27$ drivers completed the study in Canada. The following sections provide the results of the study for each group of outcomes. (NOTE: To avoid breaking up the text with the large number of data tables, we opted to locate all of the data tables the end of the report.)

6.1 Copilot® (PERCLOS), SafeTRAC®, and AP+® truck outcomes: Canada Phase

Table 2 provides the unweighted analyses of changes in mean values for NO FEEDBACK and FEEDBACK conditions (i.e., paired comparisons), for changes in standard deviations (variability) for PERCLOS at night (recorded by the Copilot®) and for the SafeTRAC® Driver Alertness score, as well as for the changes in standard deviations for lateral distance, for steering wheel movements, and for front wheel movements (all recorded directly by the AP+®). The bottom portion of the Table also shows the results of analyses on AP+® variables for truck movement (speed, engine rotation, X and Y acceleration), and ambient light level. Table 3 summarizes the mixed model (doubly weighted) analyses for the same parameters as Table 2. Table 4 provides the unweighted analyses of changes in *median* values for these same variables, as well as for the changes in the *interquartile ranges* (IQR) for these variables. Table 5 summarizes the mixed model (doubly weighted) analyses for the medians and IQR for the same variables listed in Table 4. The sum of total hours during the NO FEEDBACK and FEEDBACK conditions used as weighting factors in the mixed models are contained in Table 6. The results in these Tables are discussed below in subsections.

6.1.1 Analyses of PERCLOS (from Copilot®) during night driving ≥ 30 mph

PERCLOS (percent slow eyelid closure) obtained from the Copilot® technology during night driving above 30 mph was a primary outcome variable for hypothesis testing. It was defined in the analysis plan as the “average numeric indication of drowsiness during night drive time.” In Table 2, the unweighted analysis of mean PERCLOS is presented. Four drivers were excluded from this analysis. Drivers 1 and 13 had no PERCLOS values available at night in the cleaned analysis sample. Drivers 12 and 17 only had PERCLOS values available at night in the NO FEEDBACK condition. In Table 2, the mean PERCLOS value in the NO FEEDBACK condition was 6.65. (This value differs from that found in the last row of Data Quality Table 25 [Appendix C-1] which is 6.3. This is because Data Quality Table 25 includes Drivers 12 and 17 in this computation.) The mean PERCLOS value at night during the FEEDBACK condition was 5.03. The mean difference in the unweighted mean PERCLOS values at night was -1.63 (SD = 3.85) with minimum and maximum values of -10.52 and 2.80, respectively. A paired t-test for the null hypothesis that the mean difference is zero resulted in a non-significant $p = 0.112$. The mean difference and standard deviation of differences corresponds to a standardized effect size¹⁴ of $-1.63/3.85 = 0.423$. Two-sided paired t-tests require sample sizes of at least $n = 46$ to obtain at least 80% power to reject the null hypothesis of a mean difference equal to zero (assuming a two-sided $\alpha=0.05$ test).¹⁵

As shown in Table 3, when the analysis was repeated, weighting records by record duration and weighting observations used in the summary analysis by the total number of records available, the mean difference was -1.60 (SE = 0.91) with $p = 0.094$. ***Thus, the unweighted and weighted analyses provided some evidence that drowsiness as measured by the Copilot® index of PERCLOS (i.e., slow eyelid***

¹⁴ Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Ed. Hillsdale: Lawrence Erlbaum; 1988, 8-14.

¹⁵ Elashoff JD: *nQuery Advisor Version 4.0 User's Guide*, Los Angeles, CA: Dixon Associates, 2000.

closures) during night hours was reduced under the FEEDBACK condition compared to the NO FEEDBACK condition in drivers participating in the Canada study phase. Table 4 displays the results of repeating the unweighted analysis, but instead of comparing average changes in mean PERCLOS values, the analysis is based on average changes in the median values of the PERCLOS distributions. The average medians during the NO FEEDBACK and FEEDBACK conditions were 3.88 and 3.00, respectively. The mean change (SD) in the median values was -0.88 (2.31) with minimum and maximum values of -7.0 and 3.0, respectively. The paired t-test yielded $p = 0.15$.

6.1.2 Analyses of “driver alertness” (from SafeTRAC®) during driving ≥ 30 mph

The second primary outcome variable used in hypothesis testing was obtained from the SafeTRAC® technology during all driving above 30 mph. This was the SafeTRAC® output labeled “Driver Alertness” as estimated by a proprietary algorithm involving lane tracking. The unweighted mean values under NO FEEDBACK and FEEDBACK conditions were 82.58 and 81.80 (see Table 2), respectively for a mean difference (SD) of -0.78 (1.943) and minimum and maximum values of -4.93 and 2.32, respectively. A paired t-test for the null hypothesis that the mean difference is zero had a p-value of 0.107. Since larger values indicate greater alertness, these findings were not consistent with those found for PERCLOS at night. However, the weighted estimate of the mean change in SafeTRAC alertness was -0.24 (SE = 0.47) with $p = 0.620$ (see Table 3), suggesting no systematic difference in SafeTRAC® scores for its algorithm predicted “driver alertness” variable, between the NO FEEDBACK and FEEDBACK conditions in the Canada study phase.

When the analyses were repeated using median Driver Alertness scores from SafeTRAC®, the unweighted medians were 83.78 and 82.39 (see Table 4). The mean change (SD) in the median value was -1.39 (SD = 2.12) with minimum and maximum values of -6.0 and 1.0, respectively. The paired t-test yielded $p = 0.013$. In contrast to the analysis of mean, the weighted analysis using the mixed model retained its statistical significance with $p = 0.005$ (see Table 5). ***Thus, these data appear to suggest a statistically significant decrease in the SafeTRAC® index of driver alertness during the FEEDBACK condition, compared to the NO FEEDBACK condition. This result was contrary to PERCLOS findings and contrary to our hypothesis that feedback would improve lane tracking.***

One explanation for the inconsistency between SafeTRAC® estimates of driver alertness for driving at all times, and Copilot® estimates of driver drowsiness (PERCLOS) at night, is the different time frames from which data were acquired. Although not part of the original analysis plan, analyses of the Copilot® PERCLOS and SafeTRAC® driver alertness variables contained in Tables 2-5 were repeated, restricting attention to records in which the daylight indicator showed that driving was at night. Results are summarized in Tables 7 to 10 with the total duration weighting factors summarized in Table 11. The slight decrease in SafeTRAC® driver alertness values reflected in the SafeTRAC® alertness scores (Tables 4 and 5) was obliterated when restricting analyses to nighttime driving (Tables 7 to 10). No other substantial findings emerged. ***Consequently, there was no evidence that the***

statistically significant small decrease in SafeTRAC® estimates of driver alertness during the FEEDBACK condition (relative to the NO FEEDBACK condition) was due to drowsy driving at night in the Canada study phase. It is unknown what aspect of daylight driving could have contributed to the SafeTRAC® algorithm values of slightly reduced “driver alertness” scores during daylight driving.

6.1.3 Analyses of Lane Tracking Variability (from SafeTRAC®) during driving ≥ 30 mph

The third primary outcome measure used in hypothesis testing was Lane Tracking Variability obtained from the SafeTRAC® technology during all driving above 30 mph. Two measures of variability were examined: Lateral Distance Standard Deviation (see Tables 1 and 2), and Lateral Distance Interquartile Range (see Tables 3 and 4). All statistical comparisons between FEEDBACK and NO FEEDBACK conditions were not significant. *Thus, there was no evidence of reliable differences between conditions for the unweighted (Tables 2 and 4) or weighted (Tables 3 and 5) analyses for either the standard deviation or IQR measures of lane tracking in the Canada Study Phase.*

6.1.4 Analyses of Steering Wheel and Front Wheel Movement Variability (from AP+®) during driving ≥ 30 mph

A fourth class of outcomes also evaluated relative to the primary hypothesis were steering wheel mean variability and front wheel movement variability obtained from the AP+® system during all driving above 30 mph. Although mean and median steering wheel standard deviations and interquartile ranges tended to decline in the FEEDBACK condition, the differences were not statistically significant between the NO FEEDBACK and FEEDBACK conditions (Tables 2-5). The front wheel movement variability changes were smaller and also not statistically significant. *Therefore, there were no statistically reliable differences between conditions for measures of steering wheel variability in the Canada study phase.*

It was expected that the Howard Power Center Steering® (HPCS) system, which was to be used by drivers in the FEEDBACK condition, but not in the NO FEEDBACK condition, would have significantly reduced steering variability in the former relative to the latter. Since this outcome did not occur according to the steering data provided by the AP+® system, we examined whether drivers used the HPCS system and what reactions they had to the system. Data Quality Control Table 28 and 29 (Appendix C-1) revealed that the AP+® system logged in that only 8 drivers in the Canada study phase had the HPCS system on (1 = in use) for any period of time in the FEEDBACK condition, and that 7 of these 8 drivers also had the HPCS system turned on for some portion of the time in the NO FEEDBACK condition. Consistent with the protocol, the AP+® system logged that 12 drivers had the HPCS system off (0 = not in use) throughout the NO FEEDBACK condition. These data—namely that only a third of drivers used the HPCS system in the FEEDBACK condition—are doubtful however, and likely false due to technical interface problems between HPCS system sensors and the AP+® system. Both experimenters and drivers confirmed that virtually all drivers avoided using the HPCS system in the NO FEEDBACK condition, and they used it in the FEEDBACK condition. Moreover, drivers’ responses on the *Human Factors Structured Interview Questionnaire* confirm

their adherence to proper use of the *HPCS* system (see Section 10.0). Not only did virtually all drivers indicate they used the *HPCS* system only during the FEEDBACK condition, but drivers rated the *Howard Power Center Steering*® system higher than *SafeTRAC*®, *SleepWatch*®, and *Copilot*® (PERCLOS) systems (Tables 77-79). There is no obvious explanation to reconcile the high rate of driver satisfaction with the *Howard Power Center Steering*® system relative to the low rate of driver use of the system as indicated by the *AP+*® system. We believe the problem was with the steering sensors used to inform the *AP+* system that the *HPCS* was being used. These seemed to have failed or transmitted faulty information in many instances in which the *HPCS* was actually used in the Canada study phase. We had no evidence that either the *HPCS* or *AP+* technologies were not working correctly. ***Hence, we believe the drivers' extensive reports (confirmed by experimenters) of regular use of HPCS during the FEEDBACK condition are accurate, and that HPCS functioned reliably, but there were problems transmitting reliable steering sensor data to the AP+ black box recorder in the Canada study phase.***

6.1.5 Analyses of Truck Motion Variables (from *AP+*®) during driving ≥ 30 mph

For completeness, the other *AP+*® parameters were subjected to the same analyses. These included truck motion variable (vehicle speed, engine rotation, longitudinal acceleration [X], lateral acceleration [Y]), and ambient light. Differences for these variables were not a priori hypothesized to be different between NO FEEDBACK and FEEDBACK conditions, and this was the case for all truck motion variables in the Canada study phase (Tables 2-5). Ambient light level was slightly higher in the FEEDBACK condition (Tables 2 and 3).

6.2 Psychomotor Vigilance Task (*PVT-192*) performance outcomes: Canada Phase

As described in the Methods section, drivers were provided with a portable psychomotor vigilance task (*PVT-192*) test device while on the road, to provide information on their behavioral alertness as assessed by reaction-time (RT) based vigilance performance at the midpoint and end of each driving workday. The *PVT* is a well-validated 10-minute laboratory test of behavioral alertness that is widely used to obtain an estimate of performance limits in alert and drowsy subjects. It was hypothesized that relative to the NO FEEDBACK condition, FMT FEEDBACK would reduce *PVT* performance lapses, improve median RT performance, and reduce subjective sleepiness (as measured by a visual analog scale [VAS] drivers completed at the end of each *PVT* task trial).

6.2.1 *PVT-192* performance variables

PVT results in the NO FEEDBACK and FEEDBACK conditions for day, evening and nighttime tests in the Canada study phase are summarized in Table 12. The total numbers of 10-minute *PVT* trials in the NO FEEDBACK condition during the daytime, during the evening, and at night were 98 trials, 109 trials, and 73 trials, respectively, among the 20 drivers in the cleaned analysis sample from the Canada study phase. Similarly, in the FEEDBACK condition, there were 80 trials, 84 trials, and 53 trials during the day, evening, and nighttime intervals, respectively. Not all drivers had trials in all time intervals. The patterns of available data as well as descriptive statistics for each *PVT* parameter are provided in the *PVT* descriptive Tables (see *PVT* Table 1 in Appendix D-1). From this Table it can be seen that during

the NO FEEDBACK condition, 19 of 20 drivers had at least one trial during the day, 19 of 20 had at least one trial during the evening, and 15 of 20 had at least one trial during the night. Similarly, during the FEEDBACK condition, 18 of 20 drivers had at least one trial during the day, 18 of 20 drivers had at least one trial during the evening, and 14 of 20 had at least 1 trial during the night.

With rare exception, drivers' responses on the PVT were indicative of high compliance to task instructions. Typical healthy alert adults performing a 10-minute PVT test under controlled laboratory conditions have fastest reaction times averaging between 190 ms (milliseconds) and 210 ms; median reaction times averaging between 240 ms and 255 ms; fewer than 4 lapses per 10-minute test trial (i.e., lapse defined as an RT \geq 500 ms); and fewer than 5 response errors (i.e., false starts). Table 12 (and PVT Descriptive Data Tables 1-4 in Appendix D-1) revealed that the Canada volunteer drivers performed within these normative limits approximately 80% of the time. Comparably high compliance data on the PVT were obtained from the U.S. volunteer drivers (see Appendix D-2).

The analysis plan indicated that the total number of vigilance lapses, median response time, and subjective sleepiness by visual analog scale at the end of each PVT trial were to be considered the primary PVT outcome variables. The remaining variables were analyzed as secondary outcome variables.

6.2.1.1 Mixed model analyses of PVT-192 responses: Lapses (RTs \geq 500 ms)

The intraclass correlation for PVT raw lapses was 0.473 ($p = 0.0018$), which indicates that 47.3% of the variance among the number of vigilance lapses was attributable to systematic differences among Canadian drivers after accounting for time-of-day effect and fatigue management condition effect. As shown in Table 12, the interaction between time-of-day and fatigue management condition (NO FEEDBACK vs FEEDBACK) was statistically significant ($F = 5.78$, $df = 2, 24$, $p = 0.009$). Thus, the differences in the mean number of lapses between PVT trials during the NO FEEDBACK condition and the FEEDBACK condition significantly varied between trials during the day, evening, and night. During daytime trials, the model predicted mean number of lapses per trial during the NO FEEDBACK and FEEDBACK conditions was 1.95 and 3.89, respectively ($t = 4.49$, $df = 16$, $p = 0.0004$). During evening trials, the model predicted numbers of lapses per trial during the NO FEEDBACK and FEEDBACK conditions were 1.66 and 2.30, respectively ($t = 2.10$, $df = 16$, $p = 0.052$). In contrast, during night trials the model predicted fewer lapses (albeit not statistically significant) in the FEEDBACK condition compared to the NO FEEDBACK condition (2.51 vs. 2.34; $t = -1.02$, $df = 10$, $p = 0.332$).

Thus, total numbers of PVT lapses significantly increased during the daytime and evening period in the FEEDBACK condition relative to the NO FEEDBACK condition. However, there were no differences between FEEDBACK and NO FEEDBACK conditions in total lapses per trial at night. That lapses (long reaction times) on the PVT performance test were found to occur more at night is consistent with extensive data showing that performance on the PVT is more likely to be reduced when drowsiness is high, and this is

more likely at night than during daytime or evening. The surprising finding in the Canadian study phase that lapses were elevated in the daytime and evening in the FEEDBACK condition (relative to the NO FEEDBACK condition) is consistent with the SafeTRAC® “driver alertness” results reported above (i.e., a statistically significant but small decrease in SafeTRAC® estimates of driver alertness during the FEEDBACK condition—relative to the NO FEEDBACK condition—that was not due to any difference at night time).

6.2.1.2 Mixed model analyses of PVT-192 responses: Median reaction times

The intraclass correlation for PVT median response time was very large (ICC = 0.701; $p = 0.001$). Thus, repeated assessments of median response times within drivers tended to be very similar. As with total raw lapses, differences between the NO FEEDBACK and FEEDBACK conditions varied by time-of-day ($F = 3.38$, $df = 2, 24$, $p = 0.051$). During daytime trials, the model predicted PVT median response time during the NO FEEDBACK and FEEDBACK conditions were 246 ms and 257 ms, respectively ($t = 3.54$, $df = 16$, $p = 0.003$). During evening trials, the expected median response times during the NO FEEDBACK and FEEDBACK conditions were 245 ms and 254 ms, respectively ($t = 2.98$, $df = 16$, $p = 0.009$). Similar to total lapses, during PVT trials at night, the model predicted median response time was slightly lower in the FEEDBACK condition compared to the NO FEEDBACK condition but the difference was not statistically significant (256 ms vs. 255 ms; $t = -0.19$, $df = 10$, $p = 0.851$). Thus, Canada study phase results for PVT median response time were consistent with those found for PVT total raw lapses. There appeared to be a significant worsening of performance during the FEEDBACK condition during day and evening hours with a very slight but not statistically significant benefit to performance during night hours.

6.2.1.3 Mixed model analyses of PVT-192 responses: Post-PVT sleepiness rating

The intraclass correlations for the subjective post-PVT sleepiness visual analog (VAS) ratings was smaller than for PVT lapses and median response time, but was still statistically significant (ICC = 0.289; $p = 0.003$). Again, significant interaction was observed ($F = 3.72$, $df = 2, 24$, $p = 0.039$). Table 12 reveals that the pattern of the interaction was somewhat different for this subjective measure of sleepiness compared to the objective performance measures described above. There was no significant difference in expected values between the NO FEEDBACK and FEEDBACK conditions during the daytime trials (5.57 vs. 5.92; $t = 0.22$, $df = 16$, $p = 0.826$) and during the evening trials (6.56 vs. 6.16; $t = -0.52$, $df = 16$, $p = 0.608$). However, during the nighttime trials, the expected subjective sleepiness was significantly higher during the NO FEEDBACK condition compared to the FEEDBACK condition (7.56 vs. 6.18; $t = -3.20$, $df = 10$, $p = 0.009$). This finding of reduced subjective sleepiness observed for the post-PVT test sleepiness VAS rating was also observed for the pre-PVT subjective sleepiness rating (see bottom of Table 12). Thus, in terms of sleepiness ratings taken before and after PVT-192 test trials in the Canada study phase, the FEEDBACK condition appeared to reduce subjective sleepiness at night, relative to the NO FEEDBACK condition. This finding is consistent with

Copilot® data on PERCLOS showing that there was less subjective sleepiness at night during the FEEDBACK condition.

6.2.1.4 Mixed model analyses of PVT-192 responses: Secondary PVT Outcomes

Results for the secondary PVT outcomes (fastest 10% RTs; slowest 10% RTs; response errors, etc.) are summarized in Table 12. Results from the secondary objective measures were generally similar to those observed for the primary PVT outcome variables.

6.3 SleepWatch® (Actigraphy) and Sleep Management Model outcomes: Canada Phase

It was hypothesized that FMT FEEDBACK would result in objectively more sleep (actigraph determined). Table 13 provides the results of the mixed model ANOVA comparisons between the NO FEEDBACK condition and FEEDBACK condition for *SleepWatch®* (actigraphy) and *Sleep Management Model* variables. Random effects including intraclass correlations are summarized in Table 14. ICC values adjusted for feedback condition were generally large and statistically significant for all actigraphy outcomes demonstrating consistency within driver over time. ***Within the Canada study phase, none of the actigraphy outcomes demonstrated systematic changes between the NO FEEDBACK and FEEDBACK conditions (Table 13). However, more thorough analyses later in this report, in which actigraph-defined sleep data were combined across Canada and U.S. study phases and divided into Workday and Non-Workday periods, showed marked effects of FMT FEEDBACK on Non-Workday sleep durations (see Section 9.0).***

6.4 Daily diary outcomes: Canada Phase

Drivers were provided a daily diary (see Appendix B-1) to record driving conditions (weather, slow traffic, hilly roads, crosswinds, waiting); work activities (loading and unloading, deliveries, etc.); rest breaks and naps; days off; reactions to FMT devices; and day and night activities (work, rest, and sleep). Daily Diary data Tables 1 to 25 (see Appendix E-1) provide per driver quantitative summaries of the diary data for the 20 drivers in the cleaned analysis sample for the Canadian study phase (see Appendix E-2 for comparable diary data from the U.S. study phase).

Three types of Daily Diary variables were summarized. Data were tabulated a number of ways, according to type of variable. The first was the proportion of days in which at least one event of a specific type was reported (e.g., a long delay in traffic). Proportions were summarized by FMT condition (FEEDBACK vs NO FEEDBACK). The second type of variable was the number of events per day. The descriptive diary Tables summarize the distributions over days for each driver separately for the NO FEEDBACK and FEEDBACK conditions. The third type of variable was the cumulative duration for the events summarized by frequency per day. These are also summarized in the Diary Tables (Appendix E-1 and E-2).

Descriptive analyses comparing the NO FEEDBACK condition to the FEEDBACK condition were performed for the mean and median cumulative duration variables (Table 15) and for the mean and median frequency per day variables (Table 16). In general,

systematic differences between conditions did not emerge with the noteworthy trend for the mean cumulative daily in-vehicle nap duration, which increased from 1.58 hours per day in the NO FEEDBACK condition to 1.96 hours per day in the FEEDBACK condition ($p = 0.117$). The mean (SD) difference was 0.37 (0.99), reflecting a standardized effect size of 0.39. A sample size of at least $n = 54$ would be necessary to achieve at least 80% power to reject the null hypothesis of no change in mean daily cumulative in vehicle nap duration assuming a two-sided $\alpha = 0.05$ test. A one-sided test for research hypothesis of increased mean cumulative sleep/nap durations would require $n = 43$ drivers. *Thus, it appears that there was no evidence from drivers' daily diaries to support the hypothesis that FMT FEEDBACK resulted in increased sleep time relative to NO FEEDBACK.* Again, however, later analyses on actigraphically-defined sleep durations from both study phases revealed a clear positive effect of FMT FEEDBACK on Non-Workday sleep duration (see Section 9.0).