# Driver Drowsiness Mitigation through Real-time Drowsiness Detection

**Birsen Donmez*, Suzan Ayas**

*Professor, University of Toronto

**Canada Research Chair** in Human Factors and Transportation

October 14, 2025

Mechanical & Industrial Engineering
UNIVERSITY OF TORONTO

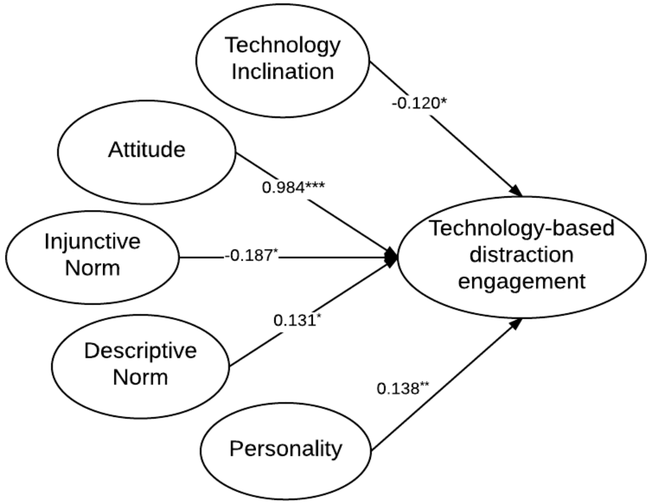
Human Factors & Applied Statistics Laboratory

Human Factors &
Applied Statistics
Laboratory


https://hfast.mie.utoronto.ca

# Our research in driving

- **Driver inattention/distraction**
- Driver experience and anticipatory driving skills
- ADAS HMI design, knowledge, training
- AVs: eHMIs, public acceptance
- VRU safety
- Commercial truck driver training; fleet safety
- …




**US pedestrian deaths are soaring. Is it time to ban right turns on red lights?**

Nearly fifty years after the federal government pushed for looser rules, cities across the country are considering a change

# Driver distraction mitigation

## Safety implications of providing real-time feedback to distracted drivers

Birsen Donmez[a], Linda Ng Boyle[a,b,*], John D. Lee[a,b]

## Mitigating driver distraction with retrospective and concurrent feedback

Birsen Donmez, Linda Ng Boyle*, John D. Lee

ELSEVIER journal homepage: www.elsevier.com/locate/aap

## Exploring the Behaviour Change Wheel and the Theoretical Domains Framework in interventions for mobile phone driver distraction: A scoping review

zadeh Nooshabadi[a],

ELSEVIER journal homepage: www.elsevier.com/locate/aap

Full length article

## Designing feedback to mitigate teen distracted driving: A social norms approach

Maryam Merrikhpour, Birse

*University of Toronto, Department of Mechanical*

## Mitigating Teen Driver Distraction: In-Vehicle Feedback Based on Peer Social Norms

**Birsen Donmez**, **Maryam Merrikhpour**, and **Mehdi Hoseinzadeh Nooshabadi**, University of Toronto, ON, Canada

Accident Analysis and Prevention 91 (2016) 166–174

ELSEVIER Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

## What drives technology-based distractions? A structural equation model on social-psychological factors of technology-based driver distraction engagement

Huei-Yen Winnie Chen, Birsen Donmez*

*University of Toronto, Department of Mechanical and Industrial Engineering, Canada*

## Voluntary- and Involuntary-Distraction Engagement: An Exploratory Study of Individual Differences

**Huei-Yen Winnie Chen**, University at Buffalo, Buffalo, New York, **Liberty Hoekstra-Atwood**, and **Birsen Donmez**, University of Toronto, Toronto, Canada

ELSEVIER Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

## Cell phone-related driver distraction: Habits predict behavior over and above the theory of planned behavior variables

Basar Demir[a,d], Junmin Du[b], Braden Joseph Hansma[a], Huei-Yen Winnie Chen[c], Haoshu Gu[b], Birsen Donmez[a,*]

# Scoping Reviews

1. Drowsiness scoping review- April 2022

**Drowsiness Mitigation Through Driver State Monitoring Systems: A Scoping Review**

Human Factors
2024, Vol. 66(9) 2218–2243

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00187208231208523
journals.sagepub.com/home/hfs
Sage

**Suzan Ayas[1], Birsen Donmez[1], and Xing Tang[1,2]**

Open Access

2. Distraction scoping review- May 2024: in prep.

# Eligibility Criteria

## Inclusion criteria

- in-vehicle interventions for driver **drowsiness** or **distraction**
- a system
  - uses a rule or algorithm to detect driver state
  - initiates an intervention

## Exclusion criteria

- **not in-vehicle technology**
  - road strips
  - regulations.
- **not use a DMS**
  - driving apps for phone

# Search and Screening Process

## **Drowsiness** - April 2022

3260 articles identified

- IEEE Explore (n=1244)
- Scopus (n=2043)
- ACM DL (n=160)
- SAE (n=52)
- TRID (n=431)

740 full-text articles screened

20 eligible drowsiness studies

## **Distraction** - May 2024

5694 articles identified

- IEEE Explore (n=1380)
- Scopus (n=2419)
- ACM DL (n=192)
- SAE (n=77)
- TRID (n=513)

Research Rabbit (n=1404)

1067 full-text articles screened

28 eligible distraction studies

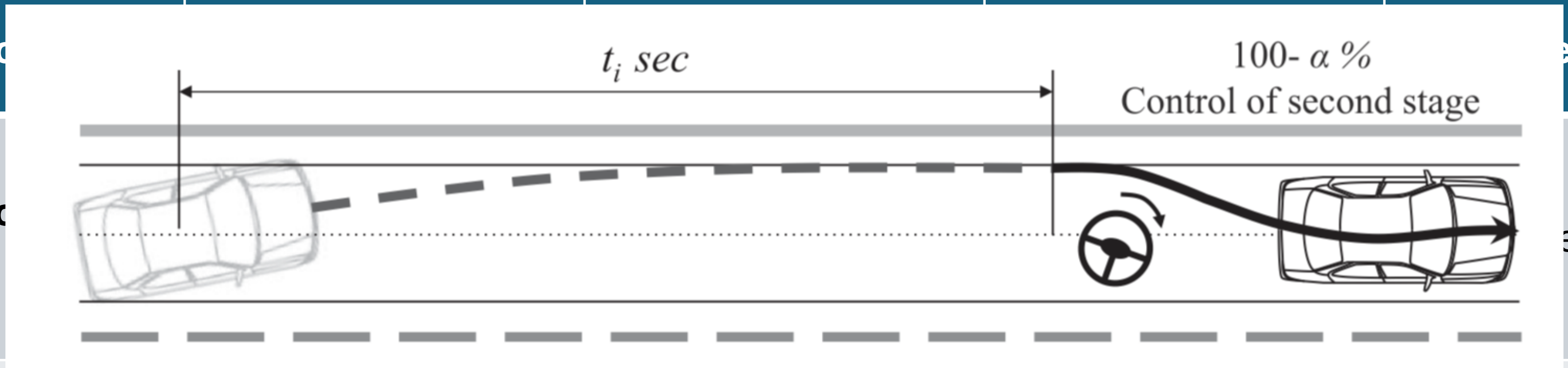
Intervention Design/Effectiveness Understudied!

# Intervention Types

| Intervention Type | Goal | How it Works (Brief) | Driver Action Expected | Example(s) |
|---|---|---|---|---|
| **1. Feedback Systems** | Inform driver of current state | Visual displays (e.g., scores, icons) | Increase awareness; self-correct | Drowsiness score |
| **2. Warning Systems** | Alert to risk; prompt corrective action | Salient alerts (visual, auditory, haptic); often staged | Immediate response to mitigate risk | Beep/flash; seat vibration |
| **3. Automation / Intervening Systems** | Vehicle actively intervenes; modifies behavior | Modifies ADAS; takes control (steering/braking) | Passive acceptance or resume full control | ACC less aggressive; lane keeping support |

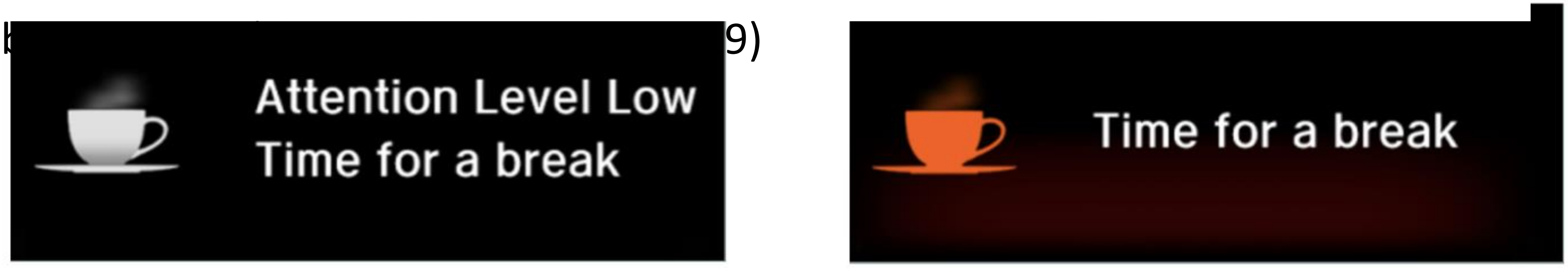
# Intervention Types

| Interventio | | | | e(s) |
|---|---|---|---|---|
| **1. Feedbac Systems** | | | | ess score |
| | | | | eat |
| **3. A Inte Sys** | | | | |



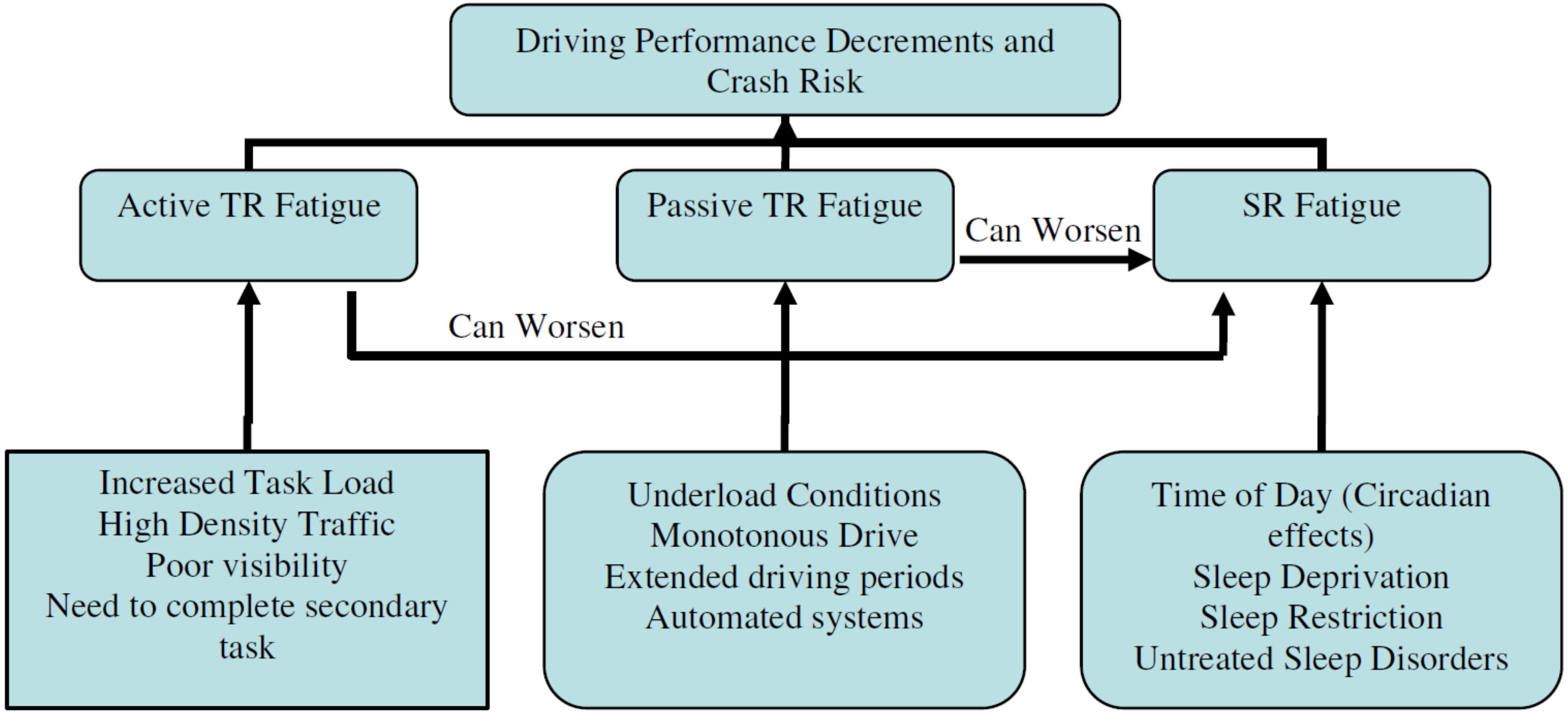


Feedb 9)




Warning Display (Nishigaki & Shirakata, 2019)

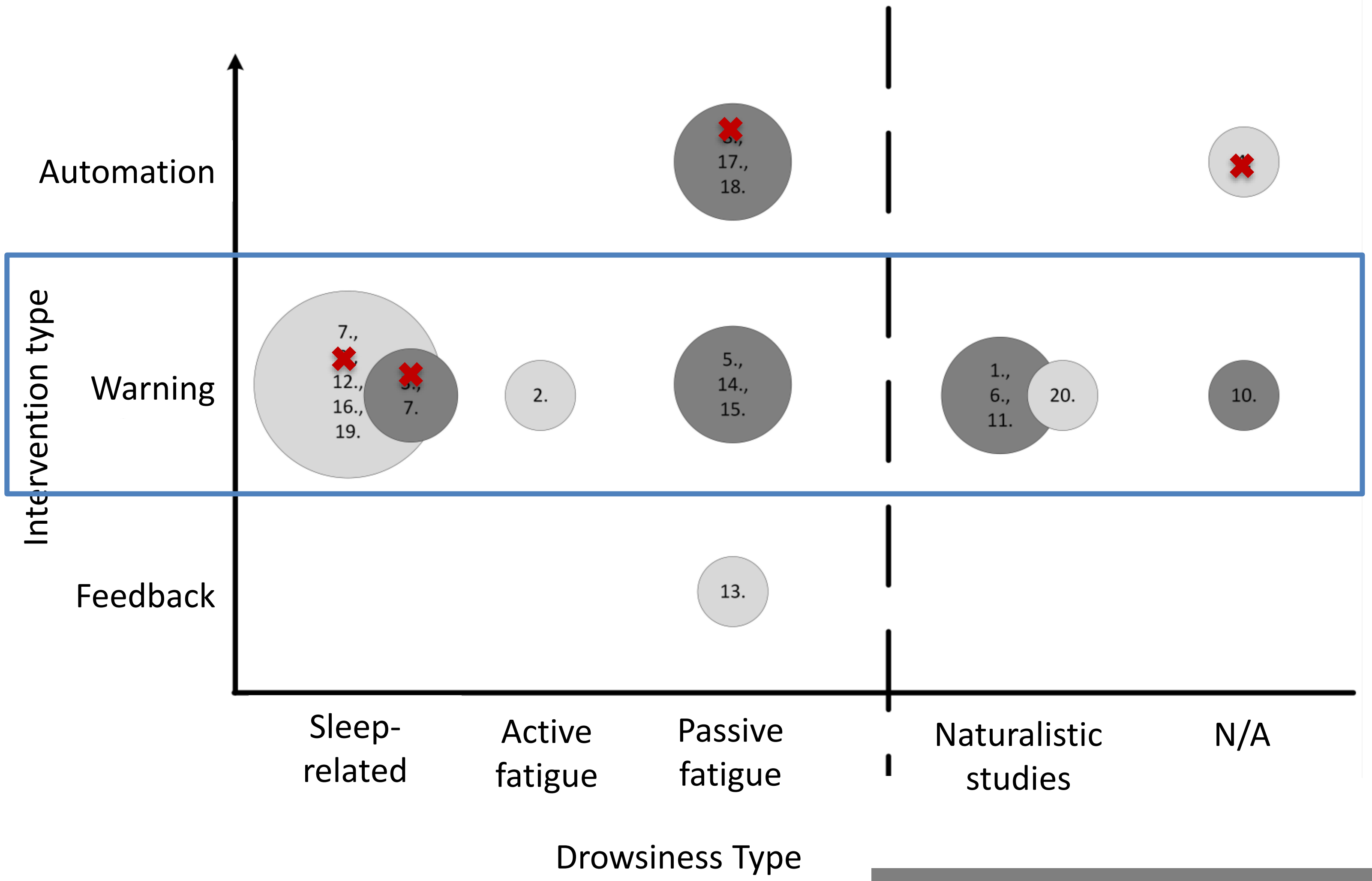
# Drowsiness Types




May and Baldwin, 2009

**Fig. 1.** A model of fatigue.

# Findings and Gaps




*Dark gray: multistage intervention e.g., caution and urgent levels*

- Generally effective in reducing sleep levels and improving driving measures **in the short run**
- Effectiveness for longer drives and long-term benefits unclear for the general population
  - **Need to go beyond warnings**

**#5.**
**Feedback + warning & indirect detection & passive fatigue**
**Monetarily Incentivized to continue driving**

TRANSPORTATION HUMAN FACTORS, 2(3), 229–246


# The Influence of Impairment Feedback on Driver Behavior: A Simulator Study

Stephen H. Fairclough
*The HUSAT Research Institute*
*Loughborough University*

Wim van Winsum
*TNO Human Factors Institute*
*Soesterberg, The Netherlands*

A number of technological countermeasures have been proposed to reduce the incidence of accidents due to driver impairment, that is, the degradation of driving performance due to the influence of fatigue, alcohol, drugs, distraction, and emotional stress. Few studies have been concerned with those driver–system interaction issues underlying this category of technology. Eighteen male participants took part in a repeated measure design in which they performed simulated journeys with and without impairment feedback. The diagnosis and assessment of impairment were based on the quality of vehicular control. Impairment feedback was presented in the form of two interface designs, one providing 3 levels of feedback and another capable of 9 levels of feedback. The results indicate that impairment feedback counteracts the characteristic degradation of driving performance due to time-on-task with respect to vehicular control. However, the presence of feedback (in either form) failed to influence participants' decision to discontinue the journey. In addition, impairment feedback failed to significantly influence psychophysiological effort, subjective fatigue, or subjective mental workload. The implications of these findings for future research and development are discussed.

# #7.
# Warning & indirect detection & passive fatigue

## Evaluating driver drowsiness countermeasures

John G. Gaspar, Timothy L. Brown, Chris W. Schwarz, John D. Lee, Julie Kang & James S. Higgins

**To cite this article:** John G. Gaspar, Timothy L. Brown, Chris W. Schwarz, John D. Lee, Julie Kang & James S. Higgins (2017) Evaluating driver drowsiness countermeasures, Traffic Injury Prevention, 18:sup1, S58-S63, DOI: 10.1080/15389588.2017.1303140

**Table 1.** Demographic data.

| | Countermeasure condition | *N* | Male/female | A *M* (SD) | | |
|---|---|---|---|---|---|---|
| | No countermeasure | 24 | 12/12 | 26.58 (4.60) | | |
| Binary | Auditory–visual (A-V) | 8 | 4/4 | 23.34 (1.60) | | |
| | Haptic | 8 | 4/4 | 25 (2.33) | | |
| | Combined (A-V + Haptic) | 8 | 4/4 | 26.13 (4.61) | 21 | 34 |
| Staged | Auditory–visual (A-V) | 8 | 4/4 | 24.63 (3.34) | 21 | 30 |
| | Haptic | 8 | 4/4 | 24.88 (1.81) | 23 | 28 |
| | Combined (A-V + Haptic) | 8 | 4/4 | 27.63 (2.97) | 23 | 32 |


Figure 1. Visual component of countermeasures for the discrete (top) and staged (bottom) alerts.

**Conclusions**: The results indicate that simple in-vehicle countermeasures, such as an auditory–visual coffee cup icon, can reduce the frequency of drowsy lane departures in the context of relatively short drives. An important next step is to evaluate the impact of drowsiness countermeasures in the context of longer, multiple-hour drives. In these cases, it may not be possible to keep drivers awake via feedback warnings and it is important to understand whether countermeasures prompt drivers to stop to rest. The next phase of this

# #12.
# Warning & direct detection (perclos) & triggered based on driving performance

*PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 50th ANNUAL MEETING—2006*

## EVALUATION OF LANE DEPARTURE WARNINGS FOR DROWSY DRIVERS

Ksenia Kozak, Jochen Pohl, Wolfgang Birk, Jeff Greenberg, Bruce Artz, Mike Blommer, Larry Cathey, Reates Curry
Ford Motor Company, Dearborn, MI

Lane departure warning (LDW) is a driver warning system designed to reduce the number of unintended lane departures. We addressed warning effectiveness and customer acceptance when the unintended lane departures are the result of drowsy driving. Thirty-two adults who were sleep deprived for 23 hours participated in the study and drove Ford's VIRTTEX driving simulator. Four Human Machine Interfaces (HMI) for LDW were evaluated: Steering Wheel Torque, Rumble Strip Sound, Steering Wheel Vibration and Head Up Display. A yaw deviation technique was used to produce controlled lane departures in the first two hours of the drive while for the last 20 minutes driver-initiated lane departures were analyzed. The Steering Wheel Vibration HMI, accompanied by Steering Wheel Torque, was found to be the most effective HMI for LDW in a group of drowsy drivers, with faster reaction times and smaller lane excursions. The Vibration HMI was also perceived by the drowsy drivers to be acceptable and helpful.

systems. This study does not address whether a LDW system may lead to drowsy drivers driving longer and in a drowsier state than they would without such a system.

**#1.**
**Naturalistic study**
**Military sample**
**Feedback + warning**

# Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance

Eugene Aidman[a,b,*], Carolyn Chadunow[a], Kayla Johnson[a], John Reece[c,d]

[a] Defence Science and Technology Organisation (DSTO), Land Division, Edinburgh, South Australia 5111, Australia
[b] University of Sydney, School of Psychology, Sydney, New South Wales 2006, Australia
[c] RMIT University, School of Health Sciences, Bundoora, Victoria 3083, Australia
[d] Australian College of Applied Psychology, Melbourne 3000, Australia

JDS scores range from 0 to 10, with higher scores indicating increasing drowsiness. A JDS score between 0 and 4.4 (inclusive) indicates a low risk level of drowsiness. A score between 4.5 and 4.9 (inclusive) indicates medium risk and JDS scores of 5.0 and above indicate a high risk (Johns et al., 2008a). In the feedback condition, the JDS scores were displayed on a 50 mm × 80 mm monochrome LCD screen attached to the dashboard immediately to the left of the visual arch of the steering wheel. This dashboard indicator also produced auditory and visual warnings when JDS scores reached the medium or high risk range. When a drive

| | Feedback condition | |
|---|---|---|
| | On<br>Mean[a] (SE)<br>(*N*) | Off<br>Mean[a] (SE)<br>(*N*) |
| KSS | 7.59 (0.32)<br>(171) | 7.30 (0.32)<br>(282) |

The important question of why the feedback condition produced these effects, remains open. The active ingredient of our feedback condition that reduced drowsiness (JDS) and improved alertness (KSS) and performance appraisals, is yet to be explained. The most optimistic explanation would suggest a

# #6.
# Naturalistic study
# Commercial fleet.
# Feedback + fatigue management plan.

per se may not exist. The key challenge with nonfleet drivers is creating an incentive, or motivation, for drivers to take a break from driving when a fatigue event is detected.

Following the preceding point, the installation of the DMS into the commercial truck fleet was accompanied by a safety management program, including a fatigue management plan. Moreover, prior to the installation of the DMS, the transport company had very robust, sophisticated, and long-standing driver and fleet management plans, including global positioning system (GPS) monitoring of vehicles, speed management technologies, driver education, and health and safety plans.

## The relative importance of real-time in-cab and external feedback in managing fatigue in real-world commercial transport operations

Michael Fitzharris, Sara Liu, Amanda N. Stephens & Michael G. Lenné

**Methods**: Data collected from a commercially available in-vehicle camera-based driver monitoring system installed in a commercial truck fleet operating in Australia were analyzed. The real-time driver monitoring system makes continuous assessments of driver drowsiness based on eyelid position and other factors. Data were collected in a baseline period where no feedback was provided to drivers. Real-time feedback to drivers then occurred via in-cab auditory and haptic warnings, which were further enhanced by direct feedback by company management when fatigue events were det
rates and their timing of occurrence across the three tir
**Results**: Relative to no feedback being provided to drive
ings resulted in a 66% reduction in fatigue events, with
of direct feedback in addition to in-cab warnings ($p <$
duration a d occurred later in the trip, and fewer drivers

**Table 1.** Study design and time periods under observation for the freight transport commercial vehicles.

| Baseline (silent) | Intervention mode 1 (IM-1): driver feedback | Intervention mode (IM-2): driver and company feedback |
|---|---|---|
| July 21, 2011, to September 8, 2011 | September 9, 2011, to January 31, 2014 | February 1, 2014, to December 7, 2016 |

# #10
# Qualitative evaluation with participatory design sessions.
# Commercial fleet

5350 IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 23, NO. 6, JUNE 2022

## Human-Centered Design for an In-Vehicle Truck Driver Fatigue and Distraction Warning System

Tim Horberry, Christine Mulvihill, Michael Fitzharris, Brendan Lawrence, Mike Lenné, Jonny Kuo, and Darren Wood

***Abstract*—Driver fatigue and distraction are major road safety issues globally; developing driver state detection and warning technology to help reduce impairment-related incidents is a promising approach. The aim of this case study was to design an effective Human Machine Interface (HMI) for a near-market driver warning system primarily aimed at commercial truck driving. A human-centered design (HCD) process was employed for the development and evaluation. Application of HCD here was a multi-stage iterative process: a comprehensive literature review; developing a context of use description; undertaking truck driver interviews; identifying user needs and associated design requirements; conducting two design workshops; operationalizing the design; running HMI evaluation studies, and finalizing the HMI concepts. As a result of the iterative HCD process, the HMI comprises a multi-modal warning system (visual, auditory and tactile) with two levels for driver fatigue and an escalating system for driver distraction. Following this successful HCD process, further on-road evaluation work is proposed before the final version of the HMI is manufactured.**

*2. What does the company want the driver to do in response to a first stage warning, particularly for fatigue?*

Almost all participants wanted drivers to self-manage their fatigue or distraction at the first level of warning (i.e., level 1 warning stage) rather than receive a telephone call from the company to check on their status. However, there was strong agreement that the level 1 warnings should be logged by the company as a means of monitoring and managing drivers, particularly those who regularly received such warnings.

*3. If the driver is to self-manage at the level 1 warning stage, then what does the company want them to do, specifically?*

Most participants suggested that the driver's response to a level 1 warning should depend on the driver's individual needs. A one size fits all approach was deemed to be ineffective even though it was acknowledged that general guidelines as to what works best to boost safety would be given during training (e.g., taking a break when tired).

One of the managers posed the question of whether it is important to know what the driver actually does if they stop to take a break after receiving the level 1 warning. Most

# #20
# Naturalistic study
# Feedback based on HR.
# Evaluated for harsh braking events.
# Commercial Fleet.
# Low accuracy levels but effective.

The impact of heart rate-based drowsiness monitoring on adverse driving events in heavy vehicle drivers under naturalistic conditions

Alexander P. Wolkow, PhD[a], Shantha M.W. Rajaratnam, PhD[a], Vanessa Wilkinson, PhD[b], Dexter Shee[c], Angela Baker, PhD[d], Teri Lillington, B.M.B.S, F.A.F.O.E.M (RACP)[d], Peter Roest[d], Bernd Marx, BBAE[d], Carmen Chew[c], Andrew Tucker, PhD[a], Shamsul Haque, PhD[e], Alexandre Schaefer, PhD[c], Mark E. Howard, PhD[a,b,*]

[a] *Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, 18 Innovation Walk, Clayton, VIC, Australia*
[b] *Institute for Breathing and Sleep, Department of Respiratory and Sleep Medicine, Austin Health, 145 Studley Road, PO Box 5555, Heidelberg, VIC, Australia*
[c] *Neurobusiness Behavioural Laboratory, Monash University Malaysia, Building 6B, Kuala Lumpur, Malaysia*
[d] *Shell International, Carel van Bylandtlaan 16, The Hague, the Netherlands*
[e] *Jeffrey Cheah School of Medicine and Health Sciences, Department of Psychology, Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway,*

A B S T R A C T

*Objectives:* This study examined the influence of a wrist-worn heart rate drowsiness detection device on heavy vehicle driver safety and sleep and its ability to predict driving events under naturalistic conditions.
*Design:* Prospective, non-randomized trial.
*Setting:* Naturalistic driving in Malaysia.
*Participants:* Heavy vehicle drivers in Malaysia were assigned to the Device ($n$ = 25) or Control condition ($n$ = 34).
*Intervention:* Both conditions were monitored for driving events at work over 4-weeks in Phase 1, and 12-weeks in Phase 2. In Phase 1, the Device condition wore the device operated in the silent mode (i.e., no drowsiness alerts) to examine the accuracy of the device in predicting driving events. In Phase 2, the Device condition wore the device in the active mode to examine if drowsiness alerts from the device influenced the rate of driving events (compared to Phase 1).
*Measurements:* All participants were monitored for harsh braking and harsh acceleration driving events and self-reported sleep duration and sleepiness daily.
*Results:* There was a significant decrease in the rate of harsh braking events (Rate ratio = 0.48, $p < 0.05$) and a fall in subjective sleepiness ($p < 0.05$) when the device was operated in the active mode (compared to the silent mode). The device predicted when no driving events were occurring (specificity=98.81%), but had low accuracy in detecting when a driving event did occur (sensitivity=6.25%).
*Conclusions:* Including drowsiness detection devices in fatigue management programs appears to alter driver behaviour, improving safety despite the modest accuracy. Longer term studies are required to determine if this change is sustained.

same Phase 1 and Phase 2 periods (Fig. 1). If drivers in the Device condition received an alert in Phase 2, they were to follow a Fatigue Break procedure which involved: finding a safe place to park within 10 km or their nearest approved rest area; completing a Fatigue Log; recording their subjective sleepiness; contacting their haulier, and; following the fatigue management strategy advised by the hauliers. As per the hauliers' fatigue procedures, the management strategy for a Fatigue Break required drivers to take a 35 min rest break, followed by a fatigue risk assessment conducted by the haulier to determine if the driver was fit to recommence driving or if they needed to extend the rest break. In addition to completing a Fatigue Break after receiving a drowsiness alert, drivers in the Device condition were also instructed to complete a Fatigue Break if they felt fatigued at any time while driving in Phase 1 and Phase 2 of the study. Furthermore, drivers in the Control group were also instructed to complete the above Fatigue Break procedure if they sensed they were fatigued while driving at any time during Phase 1 and Phase 2.

**#19.**
**Oldest but informative study.**
**Prolonged nighttime driving on test track (~5 hours).**

**BEHAVIOURAL ADAPTATION TO FATIGUE WARNING SYSTEMS**


**Alex Vincent**
**Ian Noy**
***Andrew Laing***
Transport Canada
Canada


Paper Number 98-S2-P-21

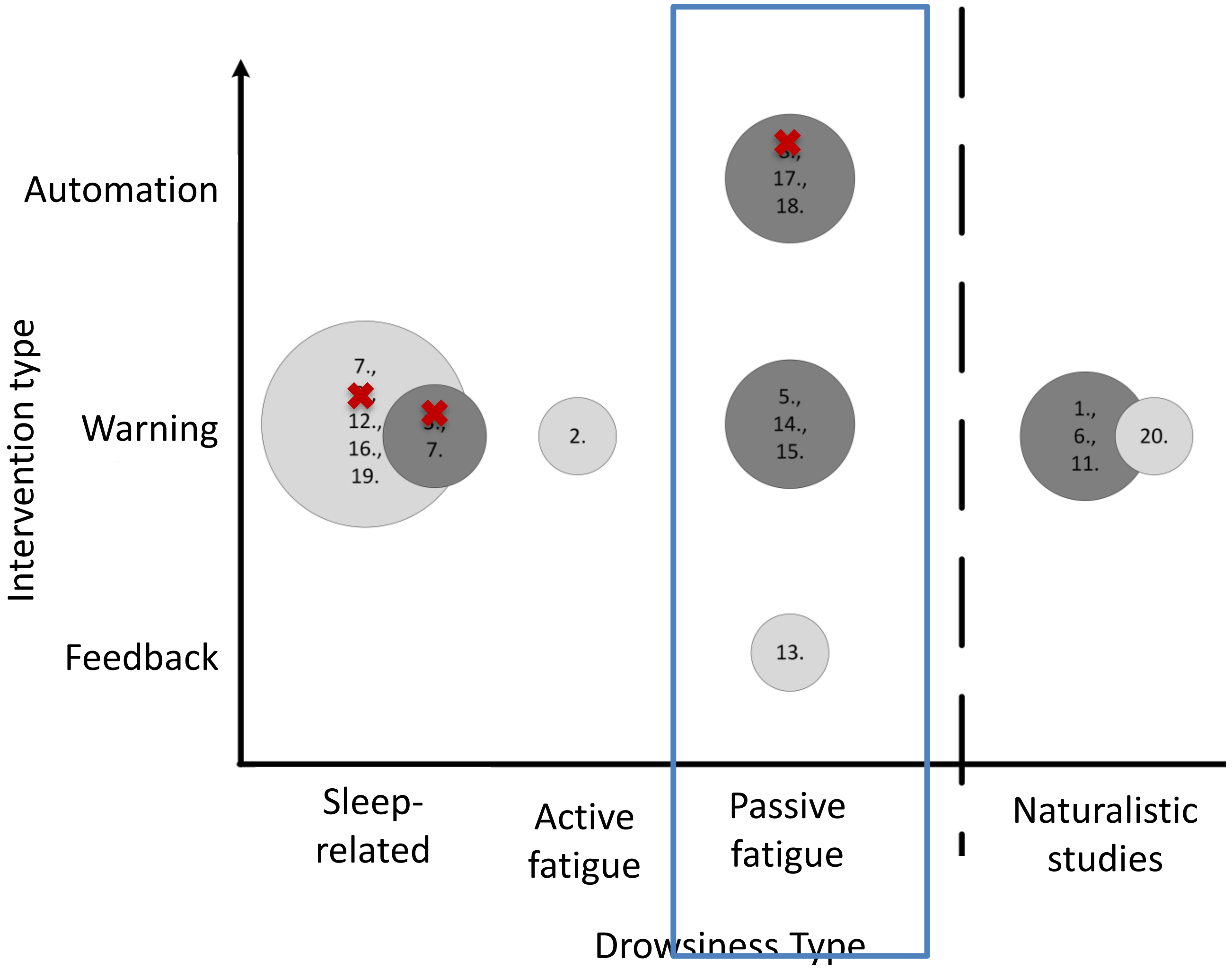

**ABSTRACT**


Driver impairment due to fatigue induced drowsiness is a significant cause of vehicle collisions. One countermeasure that is currently being implemented is Fatigue Warning Systems (FWS) to alert drivers that are drowsy. Behavioural adaptation of drivers to a FWS was evaluated in a closed track study. Thirty-two drivers completed two lengthy overnight drives, separated by one week, with half the drivers completing the second drive with an active FWS. During the drives, drivers voluntarily took breaks for as long as they liked. Behavioural results demonstrate that the FWS had no impact on objective and subjective driver fatigue, on driving time, on the number of breaks or on break duration. Results also demonstrate that 30 minute breaks are an ineffective drowsiness countermeasure . These findings suggest that a FWS as currently conceived may not contribute to reduce fatigue induced collisions.


break, drowsiness level increased monotonically and returned to pre-break levels after about 12.5 minutes. This result also holds for FWS drivers during the baseline session. However,

Future research needs to address what mechanism induces subjects to take breaks and ignore warning signals. One hypothesis is that drivers consider the signal redundant. Also, given that drivers perceive only slight decreases in their fatigue level that last a few minutes following a break, drivers are not inclined to stop and prolong the drive (cost) for a minimal improvement (benefit) in their state. This may have been a greater factor in the present study given the presence of a safety observer. Such factors reduce the effectiveness of

# Findings and Gaps




- **Automated driving interventions** for passive fatigue
  - #17, 18: two-stage control for lane keeping
  - <u>Some may make drowsiness worse</u>?
- No **cognitive task** interventions
  - Studies without DMS show improvement on driving performance and sleepiness (Atchley et al, 2014, Wörle et al, 2020)
  - DMS voice interaction: Improved lane keeping, physiological arousal, lower self-reported drowsiness (Zhang et al., 2025)
- Few (5) studies in vehicle automation setting
  - 2 ACC (#4, #17)
  - 2 SAE L2 (#13, #18)
  - 1 SAE L3 (#16)

# Two Drowsiness* Intervention Studies (Simulator)

*due to extended periods of low arousal; passive fatigue

EXP1: Manual Driving – University of Toronto
Suzan Ayas, Donmez

- Three interventions for **manual driving**:
  - **Automation**: Driving assist through SAE-L2 vehicle control
  - **Cognitive task**: 1&2back task
  - **Combined**

N=36 between subjects

EXP2: SAE L3 Driving – Ulm University
Suzan Ayas, Martin Baumann, Donmez

- Two **cognitive task** interventions for **SAE L3 driving**:
  - **Lower demand** (1 back)
  - **Higher demand** (2 back)
- No intervention condition

N=36 between subjects

# Wizard of Oz “Voice Assist” Interventions

**COGNITIVE TASK**

*“To help you stay alert, I ask you to participate in the n-back task during the next 10 minutes. During this time, you will continue to drive the car as usual so remember to keep your attention on the road while you complete the n-back task. I will guide you through it and notify you when the n-back period is over. Drive safely”*

- A modified auditory n-back task (He et al., 2019)
- 1-back: # of times two letters repeat back-to-back
  - AB**TT**DI**RR**LW -> answer: 2
- 2-back: # of times two letters repeat with one in between
  - NKR**S**E**S**PLUH -> answer: 1
- Each task repeated 5 times (1 task 30s long)
- Order counterbalanced

**SAE L2 AUTOMATION**

*“To help you stay alert, I'm activating automation for the next 10 minutes. During this time, the car will assist with steering and maintaining a safe speed. Please keep your hands on the wheel and stay attentive. I'll notify you 30 seconds before the automation is about to disengage. Drive safely.”*

- A failure event during training so that they did not expect this to be like an L3 system
- At the end of intervention, a gradual takeover request was provided

**COMBINATION**

*To help you stay alert, I'm activating automation for the next 10 minutes. During this time, the car will assist with steering and maintaining a safe speed. Please keep your hands on the wheel and stay attentive. Additionally, to further engage your mind and enhance alertness, I ask you to participate in the n-back task. I’ll guide you through the n-back task. I'll notify you 30 seconds before the automation is about to disengage. Drive safely.*

# Drowsiness Ratings: Observer and Subjective (current)

| Rating | Drowsiness level | Indicators (for Observer Coding) |
| --- | --- | --- |
| 1 | Alert | Appearance of alertness present; normal facial tone; normal fast eye blinks; short ordinary glances; occasional body movements/gestures |
| 2 | Slightly drowsy | **Still sufficiently alert; less sharp / alert looks;** longer glances; slower eye blinks; first mannerisms as rubbing face/eyes, scratching, facial contortions, moving restlessly in the seat |
| 3 | Moderately drowsy | **Eye-lid closures (1-2s);** mannerisms; slower eye-lid closures; decreasing facial tone; glassy eyes; staring at fixed position |
| 4 | Very drowsy | **Eyelid closures (2-3s);** eyes rolling upward / sideways; no proper focused eyes; decreased facial tone; lack of apparent activity; large isolated or punctuating movements |
| 5 | Extremely drowsy | Eyelid closures (4s or more); falling asleep; longer periods of lack of activity; movements when transition in and out of dozing |

(Wierwille & Ellsworth, 1994; Kundinger et al., 2020)

- **ORD failed for some participants**

Participant said they were very sleepy, weren’t assessed as such

No sample demographic requirements in guidelines.

- **Subjective ratings failed for some participants**

Clearly sleepy with droopy eyelids, said they were alert (young males)

- Instead of an either-or approach, using both is an option

# Drowsiness Ratings: Observer and Subjective (current)

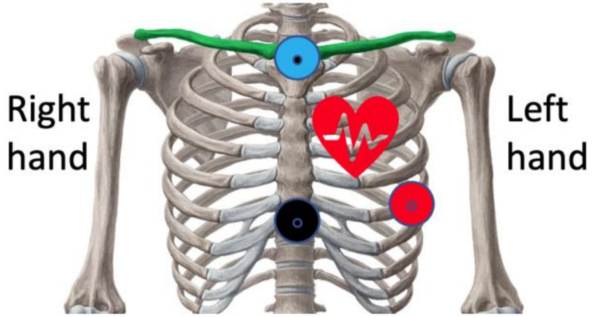
| Rating | Drowsiness level | Indicators (for Observer Coding) |
| --- | --- | --- |
| 1 | Alert | Appearance of alertness present; normal facial tone; normal fast eye blinks; short ordinary glances; occasional body movements/gestures |
| 2 | Slightly drowsy | **Still sufficiently alert; less sharp / alert looks;** longer glances; slower eye blinks; first mannerisms as rubbing face/eyes, scratching, facial contortions, moving restlessly in the seat |
| 3 | Moderately drowsy | **Eye-lid closures (1-2s);** mannerisms; slower eye-lid closures; decreasing facial tone; glassy eyes; staring at fixed position |
| 4 | Very drowsy | **Eyelid closures (2-3s);** eyes rolling upward / sideways; no proper focused eyes; decreased facial tone; lack of apparent activity; large isolated or punctuating movements |
| 5 | Extremely drowsy | Eyelid closures (4s or more); falling asleep; longer periods of lack of activity; movements when transition in and out of dozing |

(Wierwille & Ellsworth, 1994; Kundinger et al., 2020)

**POST EXPERIMENT Observer Ratings**

- Two trained independent raters per video
    - Conflicts resolved through discussions with a third tie-breaker
    - Avg. Cohen’s Kappa = 0.52, Avg. % agreement = 75%
- Rules
    - Video evaluation must be between 8am and 8pm
    - Each rater may rate a max of one hour at a time, then must take an hour of break
    - Each rater can rate max 4 hour/day
    - Each rater must rate independently (do not discuss/share ratings)

# Apparatus




Becker Meditec ECG electrodes & GSR electrodes

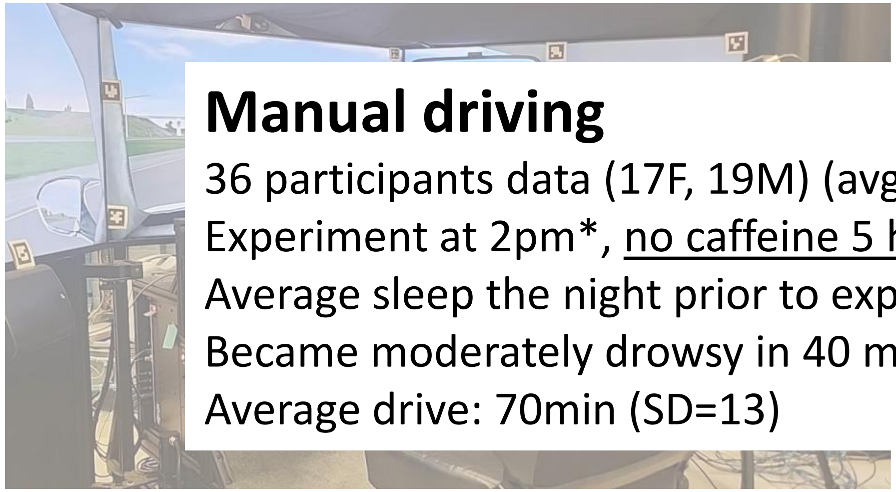
Ergoneers Dikablis 3 eye-tracker

Smart Eye Pro eye-trackers


Polar H10 ECG chestband

This presentation: Preliminary results on drowsiness ratings

# Participants



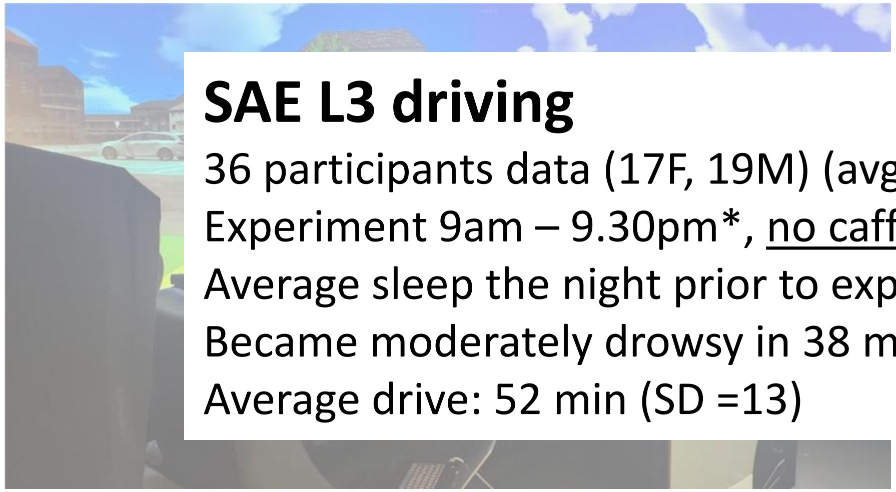
## Manual driving

36 participants data (17F, 19M) (avg. age=31, SD= 6.8)
Experiment at 2pm*, no caffeine 5 hours before
Average sleep the night prior to experiment: 8hours
Became moderately drowsy in 40 min (SD=13)
Average drive: 70min (SD=13)

## Eligibility

- A valid driver's license (Ontario G2) for the last 3 years,
- Drive ≥a few times a week **OR** have driven ≥5,000km in the past year
- No sleep related issues, no general health issues
- No corrective eyeglasses (contact lenses were allowed)
- Good written and spoken English

*No difference in time of drive



## SAE L3 driving

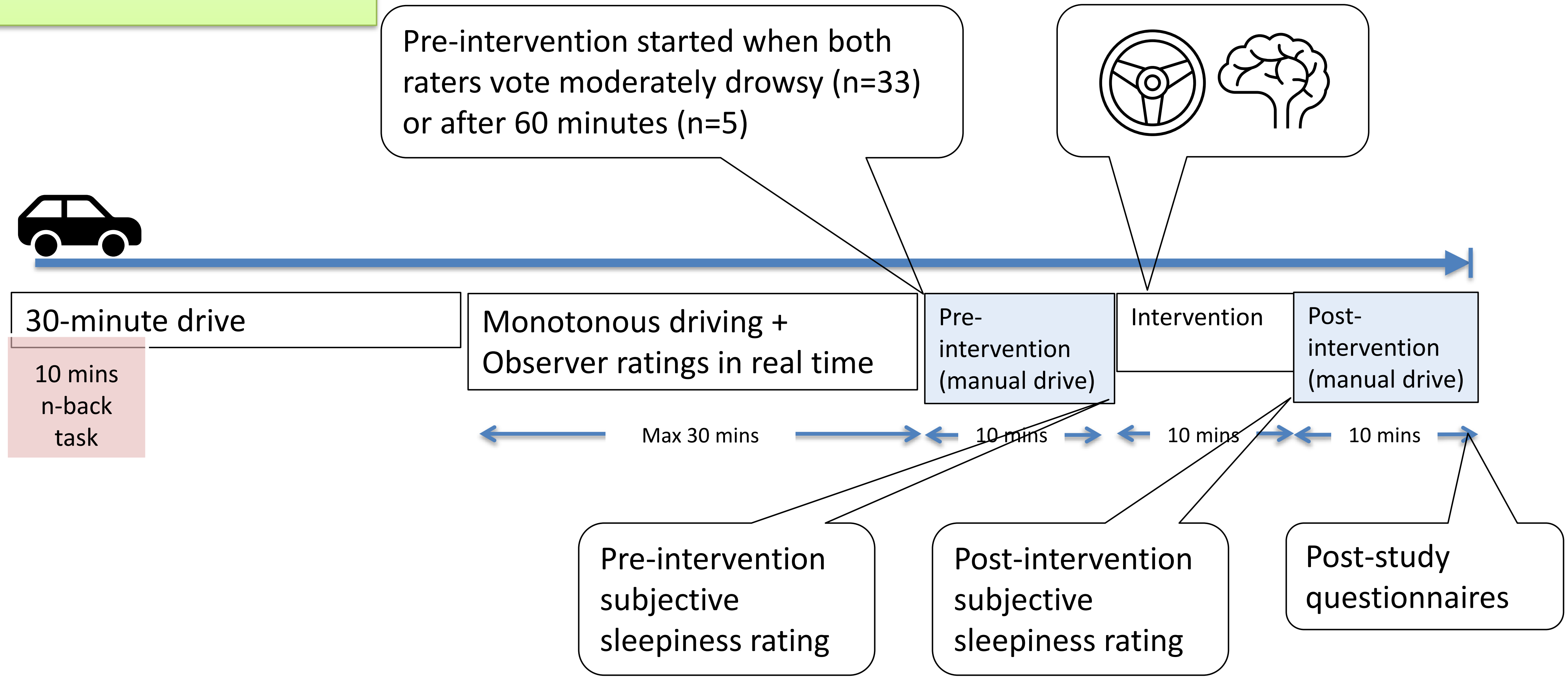
36 participants data (17F, 19M) (avg. age=27, SD= 7.0)
Experiment 9am – 9.30pm*, no caffeine 5 hours before
Average sleep the night prior to experiment: 8hours
Became moderately drowsy in 38 min (SD = 13.4)
Average drive: 52 min (SD =13)

## Eligibility

- A valid driver's license (German Class B or equivalent) for the last year,
- Drive ≥ a few times a month **OR** have driven ≥5,000km in the past year
- No sleep related issues, no general health issues
- No corrective eyeglasses (contact lenses were allowed)
- Good written and spoken English

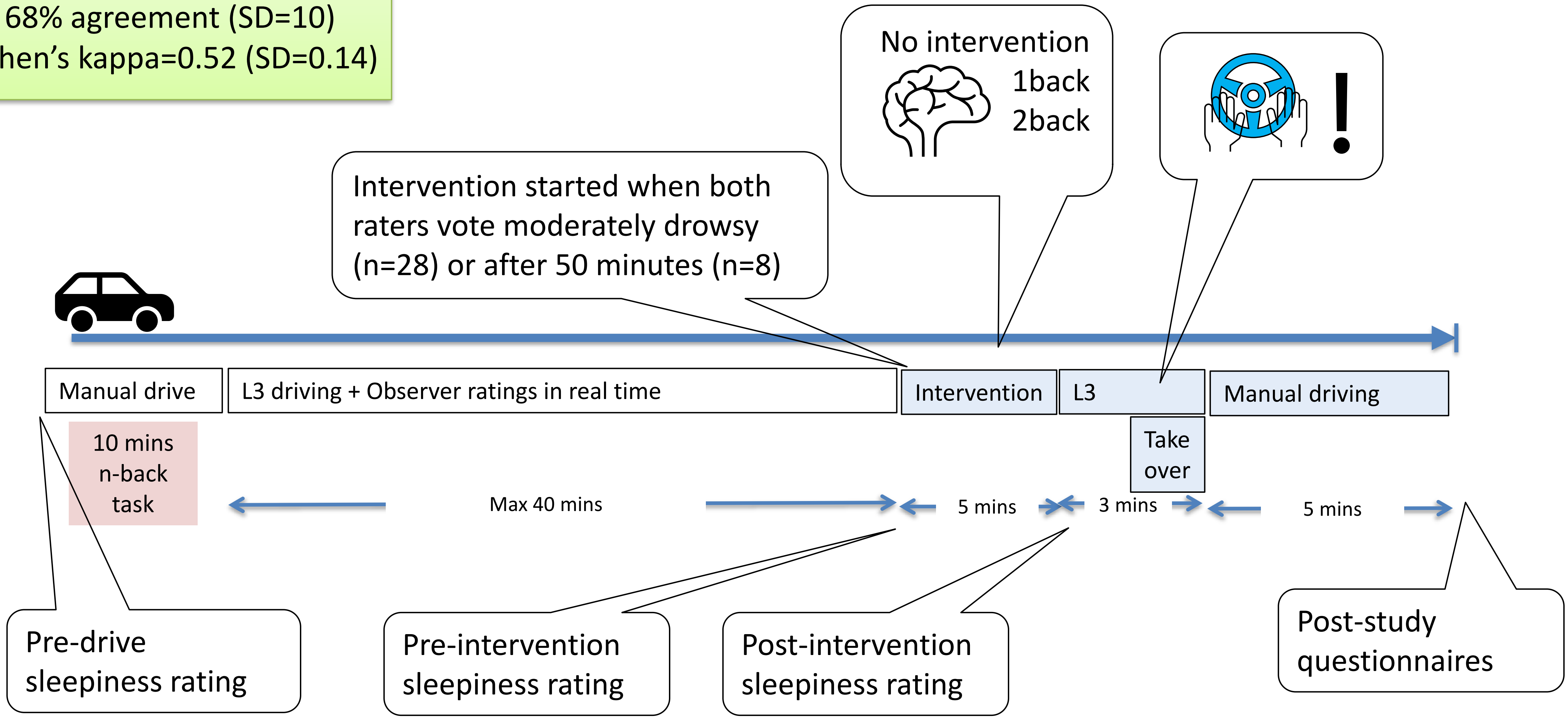
# Exp 1: Manual Driving

**Wizard of Oz DMS**
Two trained observers,
1 min video segments
70% agreement (SD=18)
Cohen's kappa=0.6 (SD=0.24)

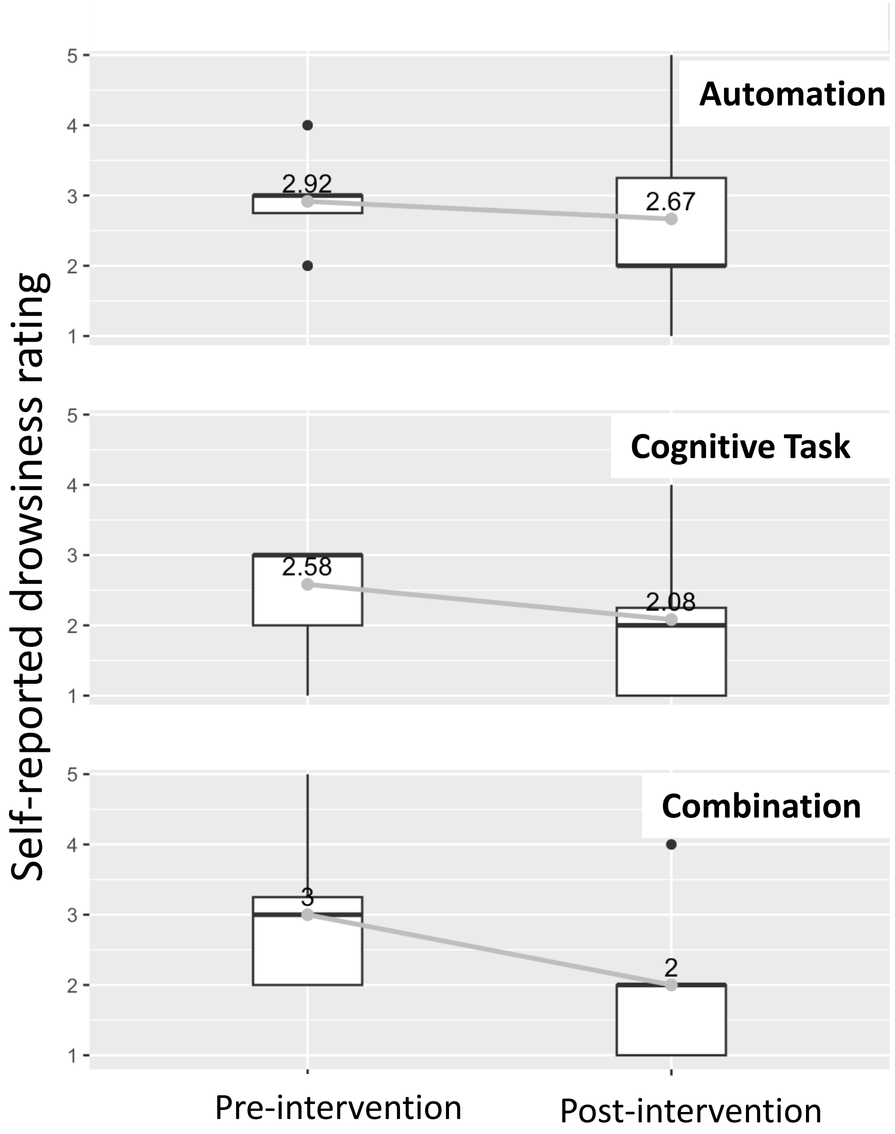


# Exp 2: SAE L3 Driving

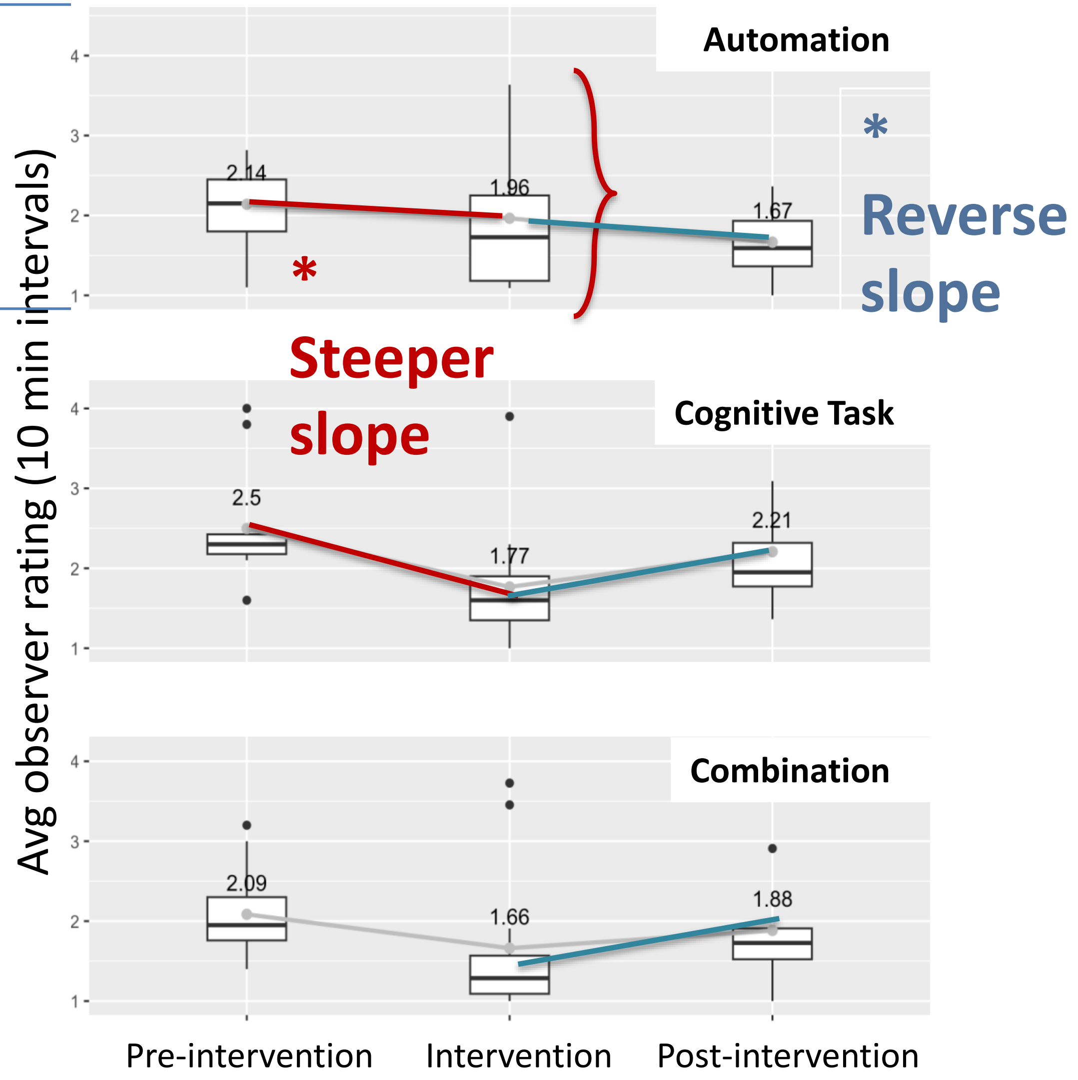
**Wizard of Oz DMS**
Two trained observers,
1 min video segments
68% agreement (SD=10)
Cohen's kappa=0.52 (SD=0.14)



# Exp 1 (manual): Interventions ↓ drowsiness ratings



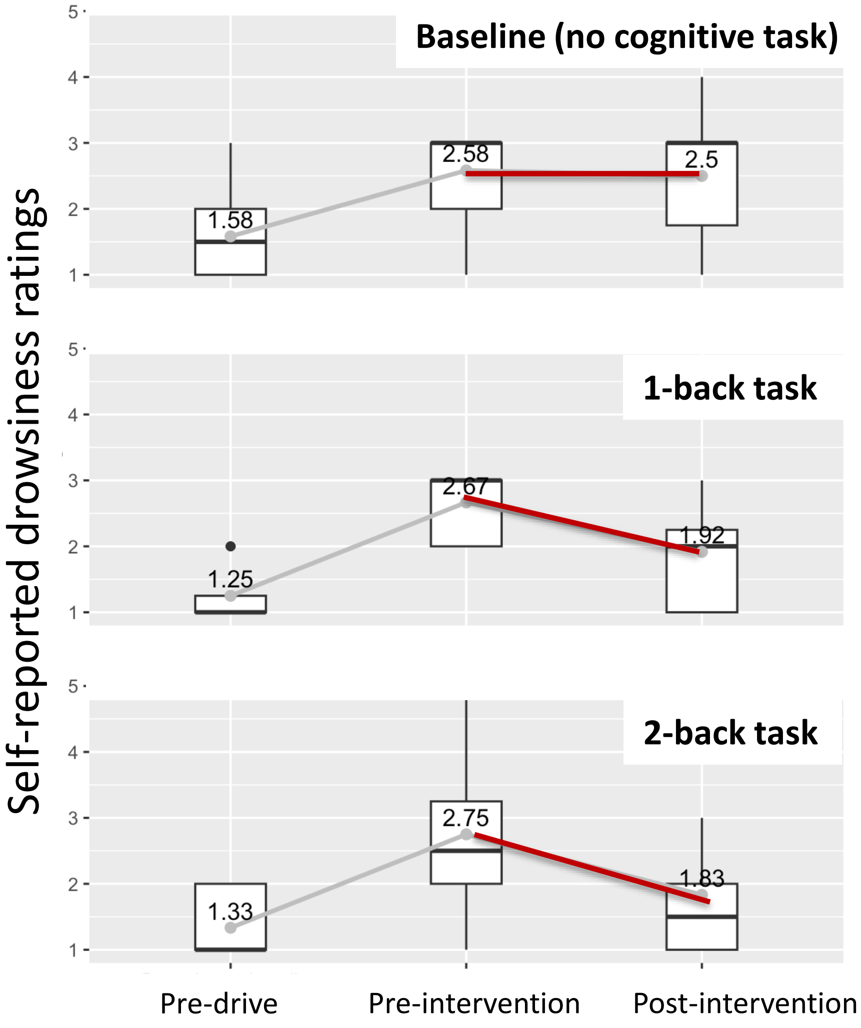

| | |
|---|---|
| 5 | Extreme |
| 4 | Very |
| 3 | Moderate |
| 2 | Slight |
| 1 | Alert |



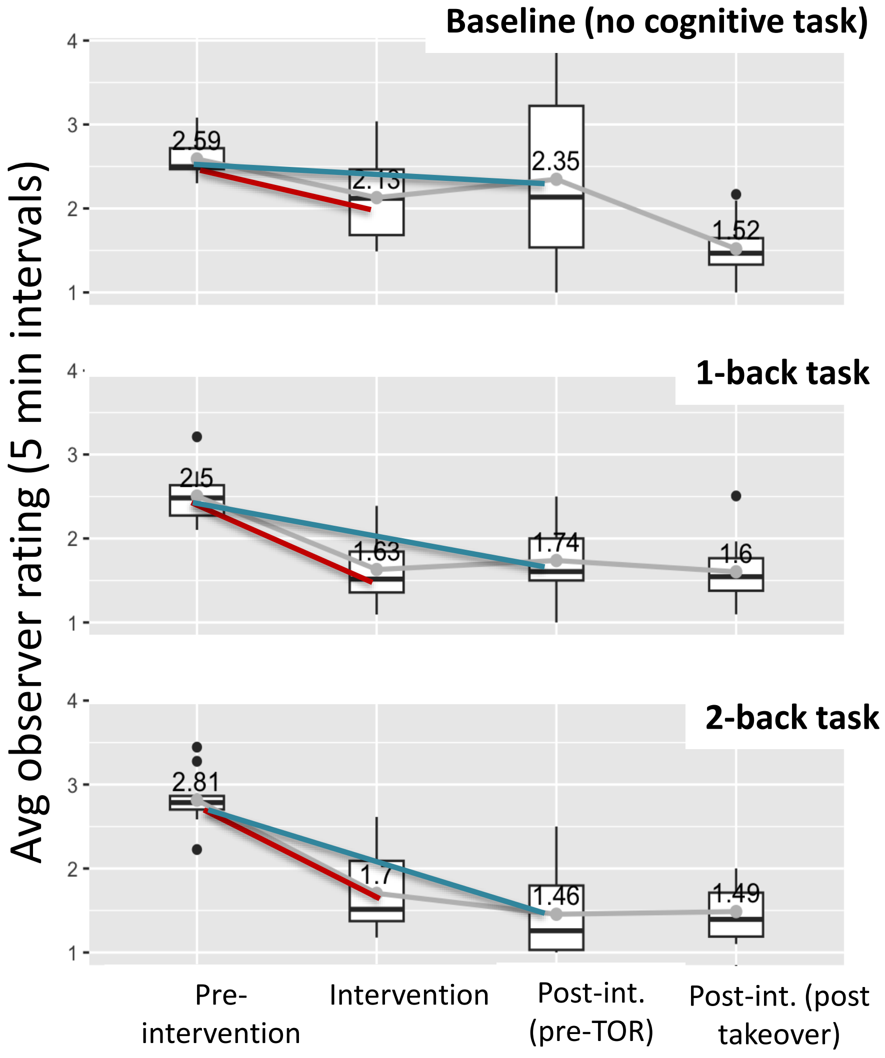

***Pre to post reduction, Δ=0.6, p=.003***
***no intervention type differences***

No differences across intervention types for post-pre values

# Exp 2 (SAE L3): Interventions ↓ ratings




| | |
|---|---|
| 5 | Extreme |
| 4 | Very |
| 3 | Moderate |
| 2 | Slight |
| 1 | Alert |




**Significantly steeper slopes for interventions**

**Significantly steeper slopes for interventions**

**Significantly steeper slopes for interventions & 2-back steeper than 1-back**

# Intervention Effectiveness (Preliminary Results)

- **Automation intervention for manual driving**:
  - some participants got rest despite the monitoring instructions
  - overreliance and misuse can be dangerous
- **Cognitive tasks**:
  - beneficial during intervention for both manual and L3 driving
  - benefits decline post-intervention
  - might have further fatigued the drivers
- **Combination** of the two did not reveal expected benefits (i.e., cognitive task compensating for potential low arousal states due to automation)
  - cognitive task alone (without driving) may not have been demanding enough
- Subjective rating prompt briefly improved ratings even for no intervention
  - even short voice interactions with DMS may momentarily improve drowsiness

# Participant Feedback

- **Automation group** (n=12)
  - 2 said automation helped them relax
  - 1 said they trusted automation because it was a simulator setting
- **Cognitive task and combination groups** (n=24):
  - 6 preferred 1-back over 2-back, another 6 preferred 2-back
  - 6 would have preferred a warning over this task intervention
  - 4 people suggested more engaging activities (e.g., podcast/music, games)
  - 4 wanted rest area suggestions
  - 4 wanted less robotic voice assistant voice
  - 3 wanted driver profile for personalized intervention
  - 3 in cognitive task group asked for driving assistance (automation)
- **3 wanted haptic and visual warnings** in addition to auditory warnings

# Takeaways

- Plethora of research on detection technology
- **Very** limited research on interventions
  – Sleep deprivation vs. active vs. passive fatigue differences underexplored
  – Warnings the “most” studied, necessary for critical levels, but are they enough?
  – L2 and L3 settings underexplored
- Both cognitive tasks (manual and L3) and automation (manual) are promising for passive fatigue with caveats
  – If drivers are required to monitor automation, they may stop doing it; automation other than L2 should be considered
  – Cognitive tasks can add to fatigue and lead to increased drowsiness post-intervention

# Corrections to Acceptance Criteria (Annex 4)

(b) The lower bound from the 90 % confidence interval of the sensitivity results shall be above 20 %. It means that 95 % of the participants statistically have more than 20 % average sensitivity, this is verified by satisfying the equation:

$$Average(Sensitivity) - 1.645 \times \frac{Standard\ Deviation(Sensitivity)}{\sqrt{Number\ of\ participants}} \geq 20\%$$

1) Not a correct interpretation of a confidence interval. The equation does NOT mean that 95% of participants statistically have more than 20% average (?) sensitivity.

- This is the confidence lower bound for the true population parameter (i.e., the mean sensitivity of system). It is about a constant value and is not about the probability distribution of sensitivity across participants, which is what should be used for the "95% of participants…" claim.

# Corrections to Acceptance Criteria (Annex 4)

(b) The lower bound from the 90 % confidence interval of the sensitivity results shall be above 20 %. It means that 95 % of the participants statistically have more than 20 % average sensitivity, this is verified by satisfying the equation:

$$Average(Sensitivity) - 1.645 \times \frac{Standard\ Deviation(Sensitivity)}{\sqrt{Number\ of\ participants}} \geq 20\%$$

2) The following comments follows if we assume that the intent of this equation is to set a lower bound for the average sensitivity estimate. Then it follows that even if one assumes that sensitivity is distributed normally (or as Gaussian as stated in the document), when we do not know the true population variance and use the sample variance instead (as is the case here), the random variable, Xbar (sample average), no longer is normally distributed – this is why we have the t-distribution. With a sample size of 10, we would have to replace 1.645 with 1.833 (from t-distribution with 9 degrees of freedom). This is less of a problem as sample size increases as t-distribution approaches z. **1.645 "incorrectly" creates an easier criterion for passing if sample size is small.**

# Corrections to Acceptance Criteria (Annex 4)

**Performance metric calculation**

The performance metrics shall be calculated as:

Sensitivity value of a participant:

$$Sensitivity = \frac{n(TP)}{n(TP) + n(FN)} \times 100\ \%$$

Average sensitivity for all participants:

$$Average(Sensitivity) = \frac{\sum Sensitivity}{Number\ of\ participants}$$

Standard Deviation (Sensitivity):

$$Stadard\ Deviation(Sensitivity) = \sqrt{\frac{\sum(Sensitivity - Average(Sensitivity))^2}{Number\ of\ participants}}$$

3) This is a question for the WG to consider. I don't claim that there is necessarily an error here.

Why not use a corrected standard deviation? In particular, why not use n-1 in the denominator? Especially for small sample sizes, uncorrected standard deviation (the current form used in the document) has larger bias.