Transportation Research Part F 48 (2017) 74–88

Contents lists available at ScienceDirect

Transportation Research Part F

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

The best way to assess visually induced motion sickness in a fixed-base driving simulator

René Reinhard [b,c], Hans M. Rutrecht [a,b,*], Patricia Hengstenberg [a], Ender Tutulmaz [a], Britta Geissler [a], Heiko Hecht [b], Axel Muttray [a]

[a] Institut für Arbeits-, Sozial- und Umweltmedizin, Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Obere Zahlbacher Str. 67, 55131 Mainz, Germany
[b] Psychologisches Institut der Johannes Gutenberg-Universität Mainz, Binger Str. 14-16, 55122 Mainz, Germany
[c] Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

ARTICLE INFO

Article history:
Received 24 March 2016
Received in revised form 15 February 2017
Accepted 16 May 2017

Keywords:
Visually induced motion sickness
Driving simulator
Fast motion sickness scale
Ordinal mixed models
Adaptation
Habituation

ABSTRACT

Objective: Driving simulator usage is becoming more widespread, yet many users still experience substantial motion sickness-like symptoms induced by optical flow, called visually induced motion sickness (VIMS). The Fast Motion sickness Scale (FMS) allows for continuous on-line assessment of VIMS. Using mixed models for ordinal data, this study investigated how to optimally analyze FMS data, and then used the resulting models to examine the development of symptoms over time in detail. Additionally, the study explored the impact of specific VIMS-inducing road elements.

Methods: Twenty-eight healthy young adults without prior simulator experience completed six courses on two days in a fixed-base driving simulator. VIMS severity was reported every minute using the FMS. Each course included two road elements designed to induce VIMS. The data was analyzed using cumulative link mixed models.

Results: The FMS data deviated clearly from a normal distribution. Treating FMS data as ordinal led to preferable models compared to models assuming interval scale. VIMS increased within each drive and over consecutive courses, but decreased between two days separated by a week (adaptation). Adaptation was attributable to less pronounced symptom increases on the second day, both within each course and between consecutive drives. VIMS increases within each drive were less pronounced during later courses of each day (habituation). Participants differed both in general symptom levels and in their progressions of VIMS over time. Additionally, VIMS-inducing road segments could be identified as leading to higher probabilities of symptom increases.

Conclusion: Mixed models analyses of FMS data from repeated VIMS measurements can benefit from taking deviations from normal distribution and interval scale into account. The gained insights into habituation and adaptation processes, as well as into the impact of specific road elements, can help in planning and conducting future driving simulator experiments.

Abbreviations: VIMS, visually induced motion sickness; FMS, fast motion sickness scale; ANOVA, analysis of variance; AIC, Akaike information criterion.

* Corresponding author at: Psychologisches Institut der Johannes Gutenberg-Universität Mainz, Binger Str. 14-16, 55122 Mainz, Germany.
E-mail addresses: rene.reinhard@itwm.fraunhofer.de (R. Reinhard), hansrutrecht@gmail.com (H.M. Rutrecht), patricia.hengstenberg@gmail.com (P. Hengstenberg), e.tutulmaz@drk-hechtsheim.de (E. Tutulmaz), britta.geissler@unimedizin-mainz.de (B. Geissler), hecht@uni-mainz.de (H. Hecht), amuttray@uni-mainz.de (A. Muttray).

http://dx.doi.org/10.1016/j.trf.2017.05.005
1369-8478/

1. Introduction

Simulators become increasingly important in industrial and scientific research, because they can provide a safe environment for complex test scenarios with high situational control and reproducibility. They enable us to gain experience with situations that are too dangerous or too infrequent outside the simulator, e.g., driving under the influence of alcohol, medication, or sleep deprivation (Muttray et al., 2013). On the other hand, the usage of simulators can adversely affect the operator's well-being, causing motion sickness-like symptoms, referred to as simulator sickness. Understanding what factors influence simulator sickness and how it develops over time can help simulator operators develop strategies that can reduce the occurrence of such undesirable side effects.

1.1. Simulator sickness and related concepts

Both moving and fixed-base simulators can cause simulator sickness in susceptible individuals. Typical symptoms include nausea, dizziness, eye strain or, in severe cases, even vomiting. When motion sickness-like symptoms are due to optic flow and without requiring physical motion, they are called visually induced motion sickness (VIMS; Keshavarz, Hecht, & Lawson, 2015). These adverse effects can last from minutes to several hours after exposure depending on the subject and the severity of the symptoms (for an overview of VIMS, see Keshavarz et al., 2015). As we conducted our experiment in a fixed-base driving simulator, we refer to the symptoms participants reported in our experiment as VIMS.

Reports about intensity and type of simulator sickness symptoms show influences of individual susceptibility including transient changes of fitness, the kind of simulation (simulator setup and simulation design), but also by the experimenter's measurement criteria (Kolasinski, 1995; Lawson, 2015). The symptoms of motion and simulator sickness worsen with increasing duration of exposure (Keshavarz & Hecht, 2011; Kolasinski, 1995; Lawther & Griffin, 1986). Both relatively constant and very rapid increases have been reported, the latter also after a certain level of discomfort had been reached (Bock & Oman, 1982; Davis, Nesbitt, & Nalivaiko, 2015; Reason & Graybiel, 1970b). Consequently, the incidence of simulator sickness differs considerably between studies, ranging from 10 to 90% (Kolasinski, 1995; Lawson, 2015).

Two mechanisms influence the development of simulator sickness over time, habituation, understood as a short-term reduction of symptoms following exposure without lasting effects, and adaptation, here seen as a long-lasting decrease of participants' susceptibility (for an overview see Keshavarz et al., 2015). Both habituation and adaptation to a nauseating environment including moving and fixed-base driving simulators have been shown to reduce symptoms of motion sickness (Golding & Stott, 1995; Howarth & Hodder, 2008; Kennedy, Stanney, & Dunlap, 2000; Mackrous, Lavallière, & Teasdale, 2014; Reason & Graybiel, 1970a; Watson, 2000). There is considerable variability between subjects (McCauley, Royal, Wylie, O'Hanlon, & Mackie, 1976). Some sensitive participants are not capable of adapting to inertial motion (Tyler & Bard, 1949). Even increased sensitivity due to repeated exposures to visual stimuli applied with a head mounted display has been reported (Howarth & Hodder, 2008).

Another influence on simulator sickness severity relates to the design of the driving scenario. The literature indicates that design decisions, such as the inclusion of sharp turns, may lead to an increase in symptoms (Stoner, Fisher, & Mollenhauer, 2011). In their review, Classen, Bewernitz, and Shechtman (2011) evaluated the impact of "Context and Environment Factors" to be "probably predictive" (p.181) of simulator sickness. However, attempting to evaluate the impact of particular road characteristics in a naturalistic driving context is difficult and previous studies often included confounded design choices, such as simultaneous variation of road curvature and visual complexity of the scene (Mourant, Rengarajan, Cox, Lin, & Jaeger, 2007; Park, Allen, Fiorentino, Rosenthal, & Cook, 2006). More importantly, the standard tool to assess VIMS and simulator sickness, the Simulator Sickness Questionnaire (SSQ; Kennedy, Lane, Berbaum, & Lilienthal, 1993), comprises 16 items. As it takes too long to administer repeatedly during a drive, it has a very low temporal resolution. Alternative approaches to motion sickness assessment have been explored in various studies both during exposure to nauseating stimuli and afterwards (Bagshaw & Stott, 1985; Bles, de Graaf, Bos, Groen, & Krol, 1997; Bock & Oman, 1982; Diels & Howarth, 2013; Garrick-Bethell, Jarchow, Hecht, & Young, 2008; Golding, Mueller, & Gresty, 2001). Out of these self-assessment scales, only the Fast Motion Sickness Scale (FMS; Garrick-Bethell et al., 2008) has been validated against the SSQ (Keshavarz & Hecht, 2011). As it consists of only a single question concerning the user's symptom level with responses ranging from 0 to 20, the FMS offers the great advantage of frequent, repeated symptom assessments. This could in principle be used to closely monitor the development of symptoms during any given test run, and allows for a cut-off criterion (e.g. a score of 15) to prevent frank sickness (Keshavarz & Hecht, 2011).

Besides monitoring the driver's status during simulator usage, this study investigates whether quick repeated assessment can also be useful in studying VIMS and simulator sickness in greater detail. The FMS' fine-grained ratings could reveal both habituation and adaptation processes, which this study investigated by monitoring the development of VIMS over the course of several days of simulator usage, with several simulated drives on each day. Additionally, the study used the FMS to evaluate the impact of specific road segments on VIMS. For these purposes, it needs an analysis strategy that can efficiently make use of dense information contained in the FMS data. Here, we propose a mixed-effects regression model based analysis strategy to address the temporal progression of VIMS in detail. This approach is explored in the next section.

1.2. Analysis of visually induced motion sickness symptoms assessed by the FMS

To model the development of FMS over the course of the experiment, we used mixed-effects regression models,[1] also called mixed models. This approach has several advantages when compared to traditional regression techniques (West et al., 2007). Firstly, mixed models allow for more flexible approaches to missing-at-random data. This can be a practical concern for oft-repeated measurements in complex experimental setups, where the alternatives are often listwise deletion or problematic imputation procedures.

Secondly, the inclusion of random effects allows for the exploration of interindividual differences. This may be of particular interest in the case of VIMS, since, as discussed above, previous research has indicated that the development of motion sickness symptoms differs considerably between participants.

Lastly, unlike the standard ANOVA, mixed models can accommodate data that are not normally distributed or even ordinal. This may very well be the case for our FMS measure (Christensen & Brockhoff, 2013). Within the medical research literature, the debate about the appropriateness of assuming interval scale for measures akin to the FMS was recently addressed by Grimby, Tennant, and Tesio (2012), following reports of erroneous inferences attributed to the application of parametric tests to analyze ordinal outcomes (e.g. Kahler, Rogausch, Brunner, & Himmel, 2008). The authors have argued that for such scales an "increase in observed raw score", in our case the observed FMS score, "will reflect a different increase in the underlying metric range," here hypothetical latent symptom severity, "depending upon the starting point" (p. 97). According to the authors, this may be constituted by an understatement of the real increase of the underlying construct by the observed increases in the raw scores within the scale's margins, and by an overstatement within its center.

An established regression framework for ordinal data are cumulative link models, which have been adapted into mixed model approaches as cumulative link mixed models (Christensen, 2015). They can be used to both evaluate the possibility of the FMS' ordinal nature and take it into account for an optimal data analysis. Christensen and Brockhoff (2013) offer an interpretation of cumulative link mixed models as mixed models of a continuous "latent" variable. This latent variable is unobservable, but only predicted by the model using the fixed and random effects. The resulting continuous value of the latent variable is then sorted into one of the categories of the ordinal scale by comparing it to a set of increasingly high thresholds. Christensen (2015) proposed that these thresholds can be conceptualized as

1. different necessary increases in the latent variable for each boundary between two categories (flexible thresholds),
2. structured, such that the necessary increases are equal for each boundary between two categories (equidistant thresholds),
3. structured so the necessary increases differ, but the needed rise is equal for pairs of high and low thresholds with the same distance from a central category (symmetric thresholds).

Restricting the thresholds to be equidistant means that the same increase in the latent variable is needed to reach the next higher category. Symmetric thresholds could be used to model Grimby et al.'s (2012) proposal, with large increases in the latent variable needed to cross from one low value to the next higher one and, due to symmetry, from one high value to the next, whereas relatively low increases would be needed when transitioning between the middle categories. Thus, cumulative link mixed models appear to be the best and most appropriate analysis for monolithic VIMS ratings.

1.3. Aims of the study

The current study sought to assess VIMS in a fixed-base driving simulator using the FMS on several measurement points across several consecutive simulated driving courses on two different days. Firstly, it aimed to explore the distribution of the FMS scores, and to evaluate which of the three threshold proposals (flexible, equidistant, or symmetric) for cumulative link mixed models best fit the FMS data. Second, it explored the impact of driving time, both within and between days of simulator usage, on simulator sickness symptoms. A special focus was on the question of how symptom trajectories changed over time, i.e., whether symptom increases on the first day and in early courses differed from increases encountered during the second day and in later courses. The mixed models approach also allowed for the detailed investigation of interindividual

[1] A mixed-effects regression model is a statistical model that includes both fixed-effect parameters, as well as random-effect parameters in order to specify the relationship between a set of predictor variables and a dependent variable (West, Welch, & Galecki, 2007). Fixed effects denote parameters used to describe the relationship of the dependent variable, e.g., FMS scores, to concepts of interest, e.g., the time already driven in the simulator. The units of fixed effects are not randomly sampled, but fixed (West et al., 2007). Random effects, by contrast, reflect factors that are thought to have been sampled from a wider population, e.g., participants. Therefore, conclusions drawn about random effects relate to the population from which the units at hand have been drawn. Random effects can either be modeled as random intercepts, or as random slopes (West et al., 2007). As random intercepts, each unit of the random effects variable, e.g., each individual participant, is associated with an individual intercept, i.e., the model still predicts the same change of symptom scores over time but at different levels for each participant. If a random slope related to a fixed effect is included in the random effects structure, each unit of the random effects variable is predicted to be differently affected by this effect, i.e., individual developments of symptoms over time are predicted for each participant. Mixed models that differ in their fixed or random effect structures can be compared using e.g. Akaike's information criteria (AIC; Akaike, 1974) to find the preferable model. This comparison evaluates whether a variable that was added or dropped in the preferred model is useful in describing the data.

differences in symptom progression. Finally, the study sought to measure the impact of specific road segments on FMS scores, specifically the impact of "bumpy road" elements that induce quick up-down motions, which had anecdotally been shown to provoke VIMS in pretests.

2. Methods

2.1. Sample description

Participants were recruited via mailing lists and advertisements at the Johannes Gutenberg-University (Psychology Department and Medical School) in Mainz. They had to possess a driving license and a minimum of 5000 kilometers of driving experience. Normal eyesight and color vision were assessed with a Titmus 2a Vision Screener (MAICO, Berlin, Germany). Due to a possible adaptation to VIMS, participants with previous experience in driving simulators or extensive experience in computer games (more than twice a week, especially racing or driving simulation games) were excluded. Participants were instructed to abstain from alcohol or caffeinated beverages before the experiment. Intensity of physical exercise in their everyday life was surveyed on a seven-point rating scale. All participants underwent a medical examination, including the measurement of blood pressure, an electrocardiography, and a medical check-up, including the ability to balance (with closed and open eyes), neurological sensitivity, and other tests by a physician of the on-site medical staff at the Universitätsmedizin Mainz.

The study was approved by the local ethics committee and adhered to the Declaration of Helsinki and its latest amendments. Participants gave written informed consent and received financial compensation for their participation. An initial sample of 34 healthy young adults participated in this study. One participant had to be excluded because of data loss, one owing to intercurrent illness, and four participants aborted due to VIMS. In total, 28 participants completed the entire experiment (age: 18–29 years, M = 23.8 years, SD = 2.5 years; 14 women).

2.2. Driving simulator

The driving simulator was a fixed-base simulator (FOERST F10P® Dr.-Ing. Reiner Foerst GmbH, Wiehl, Germany) using a 1.80 m wide and 1.39 m high projection screen for the scene view, as well as three rear-view mirrors. Participants sat in a slightly modified Ford Fiesta cockpit with an automatic gearbox and all standard controls. A built-in audio system simulated motor sounds and other relevant noises. To improve immersion, the steering wheel vibrated depending on road and weather conditions. A climate control unit (Hareus Vötsch HC0020) was used to maintain temperatures at 22 °C and humidity at 45% during the entire experiment.

2.3. Driving scenario

Four equivalent virtual environments with 23-kilometer-long roadways were constructed. Each environment was designed to take about 20 minutes of driving time when all speed limits were obeyed. Actual driving times varied between 17 and 27 minutes (M = 20.35 min, SD = 1.57 min) due to different driving styles. The environments included four distinct segments (rural, motorway, city and desert), which were simulated with appropriate street types, road signs, traffic lights, and other roadside objects (Fig. 1). All segments were designed to reflect realistic driving scenarios. Participants followed a predefined route specified by instructions given in the simulation. During the courses, two predefined weather changes happened, with possible aquaplaning in sharp turns on wet roads.

Different scenarios, weather conditions, and road situations had been explored in pretests. Preliminary tests indicated that sharp curves, changes in the linear acceleration (sharp braking), and elements with quick up-down (e.g. dips or bumpy road surface) movements provoked VIMS symptoms. This is comparable to the literature on the effects of scenario design on VIMS severity (Stoner et al., 2011). Two stretches of road with rough road surfaces that caused the car to produce vigorous vertical movements and pitch rotations ("bumpy road") were chosen for the experiment, as these segments had been the most provocative for VIMS before. They were displayed in corresponding parts of all environments, always falling between the 11th and 12th and between the 17th and 18th FMS measurement points.

2.4. Questionnaires

2.4.1. Fast Motion sickness Scale (FMS)

The FMS (Keshavarz & Hecht, 2011) is a single-item measure using a numeric response format ranging from 0 (*no nausea*) to 20 (*extreme nausea*). Participants verbally reported one FMS score approximately once a minute, prompted by both a message on the screen, and a sound file.

Scores of 15 and above resulted in instructions to slowly stop the car and end the experiment. Participants were unaware of this criterion.

Fig. 1. Simulation screenshots. Rural, bumpy road, city, desert (left to right, top to bottom).

2.4.2. Simulator Sickness Questionnaire (SSQ)

To obtain more detailed information about induced symptoms, the SSQ (Kennedy et al., 1993) was administered. This questionnaire consists of 16 symptoms (e.g. nausea, or headache) rated on a four-point descriptive graphic response format ranging from *none* to *severe*.

2.4.3. Karolinska Sleepiness Scale (KSS)

The KSS (Akerstedt & Gillberg, 1990) measures fatigue, using a single item with response options ranging from 1 (*very alert*) to 9 (*very sleepy, fighting sleep, an effort to keep awake*). Participants further had to rate if they felt able to drive unimpaired in a real situation.

2.5. Procedure

The experiment took place on two days, at least one week and at most two weeks apart, to lessen immediate training effects. All experiments started in the morning at either 8:00 or 10:00.

In the beginning of each day, participants answered pre-experimental questionnaires, including the KSS, and underwent the medical check-up and tests that controlled for possible confounding influences on VIMS. Here, breath alcohol concentrations (Draeger Alcotest 6510®, Draeger Safety AG, Lübeck, Germany) had to be below the detection limit, an urine drug screening for illegal drugs (Mahsan® Kombi/DOA6, MAHSAN Diagnostika, Reinbek, Germany) needed to indicate no signs of illegal drug intake (amphetamine, cocaine, morphine, cannabis, methamphetamine and benzodiazepine), and tested caffeine in saliva specimens had to confirm that participants had not drunken any caffeinated beverages beforehand. The median KSS score at the beginning of both days was 3 with a 90^{th} percentile of 7.

To become familiar with the simulator, participants drove a three-minute training course that was representative of the environments used during the experiment. Participants were asked to adhere to all traffic rules, especially speed limits. If they disobeyed the speed limits repeatedly, the experimenter reminded them to comply.

The four virtual environments were displayed during six drives in a randomly selected order, with two random environments being displayed twice. The same environment was never driven twice in a row or more than twice in total.

On the first day, participants drove two courses, with VIMS assessments via the FMS at 20 measurement points during the drive. Between the two drives, a break of five minutes duration was scheduled. The break was extended if participants suffered from VIMS above a FMS score of 5. In these cases, the next drive started when the FMS fell below 6. The SSQ was completed before the first drive, during the breaks and at the end of the day. If signs of VIMS were evident during the experiment, participants had to wait afterwards until a physician had made sure they were safe to leave.

The experiment's second day followed the first in its general schedule, but here participants drove four courses. VIMS was again measured at 20 measurement points during the drive using the FMS and with the SSQ before the first and after each drive.

2.6. Data analysis

Raw FMS scores were analyzed using proportional-odds cumulative link mixed models as implemented in the *clmm* function of the ordinal package (Christensen, 2015) in R (version 3.2.1). The models evaluate the 120 FMS scores provided by each participant by exploiting variance produced by the factors of experimental day (day; 2 levels), the simulated driving course (course; 2 levels on day 1, 4 levels on day 2), and a consecutive number, indicating the repetition of the FMS question posed during the drive (measurement point; 20 levels). Cumulative link mixed models first use these factors to predict values on a continuous, latent simulator sickness dimension, according to a formula exemplified in Fig. 2. The predicted values are then compared to thresholds that are also estimated from the data, to derive a predicted FMS score.

To predict the latent variable, the factors of day, course, and measurement point were incorporated as fixed effects, each specifying a systematic change in symptom levels across participants with rising levels of the factor. The full model allowed for possible interaction effects between the fixed effect factors. The random effects structure was used to predict a general symptom level for each individual participant (random intercept) and to explore individual developments of symptom severity (random slopes).

The impact of specific parts of the full model, e.g., the interaction terms, were evaluated through model comparisons between two models that differed only with regards to this aspect of the model, e.g., a model with interaction terms compared to a model without interactions. Model comparisons were conducted using Holm–Bonferroni corrected likelihood ratio (LR) tests, with the test statistic given as $LR(\mathrm{df})$. Akaike weights, $w(\mathrm{AIC})$, and evidence ratios were calculated, denoting the probability that one model is the best among a set of models given the data, and the comparison of that probability between two models (Wagenmakers & Farrell, 2004). Additionally, the individual coefficients were evaluated using 95%-confidence intervals, further illustrated with Wald-statistic *p*-values. The latter should, however, only be seen as serving a descriptive function. An overview over all models used in the comparisons can be found in Table 1.

The first aim of the current study was to explore the FMS scores themselves. For this purpose, their distribution was evaluated visually and using the Shapiro-Wilk-test, both for the raw FMS scores and for natural logarithm, square-root and inverse transformations for scores offset by 1, to avoid 0 values. To assess, whether FMS scores should be treated as ordinal data, we compared cumulative link mixed models with different threshold estimates ($\mathrm{Model0a_{flexible}}$, $\mathrm{Model0b_{equidistant}}$, and $\mathrm{Model0c_{symmetric}}$). The fixed and random effect structures of these models were all specified according to the full model (Fig. 2), albeit with different threshold constraints. It should be noted that all three threshold structures proposed for usage in cumulative link mixed models allow for zero inflation, i.e., the transition from category 0 to 1 can be estimated at a higher level than subsequent increases, allowing for a greater number of predictions in category 0 (Christensen, 2015).

$$
\begin{aligned}
\mathrm{SScontinuous_{ti}} = {} & \beta_I \\
& +\beta_D{*}D_{ti} \\
& +\beta_C{*}C_{ti} \\
& +\beta_M{*}M_{ti} \\
& +\beta_{DXC}{*}D_{ti}{*}C_{ti} \\
& +\beta_{DXM}{*}D_{ti}{*}M_{ti} \\
& +\beta_{CXM}{*}C_{ti}{*}M_{ti} \\
& +\beta_{DXCXM}{*}D_{ti}{*}C_{ti}{*}M_{ti} \\
& +u_{I|i} \\
& +u_{D|i}{*}D_{ti} \\
& +u_{C|i}{*}C_{ti} \\
& +u_{M|i}{*}M_{ti} \\
& +\varepsilon_{ti}
\end{aligned}
$$

Interactions — Fixed effects

Random intercept — Random slopes — Random effects

Fig. 2. Formula used in the full model to predict the underlying continuous variable ($\mathrm{SScontinuous_{ti}}$) for the i^{th} person ($i = 1, \ldots, 28$) and an index t ($t = 1, \ldots, 120$) of the $20 * (2 + 4)$ repeated FMS observations for the i^{th} person. Terms include an intercept (I), fixed effects for day (D), course (C) and measurement point (M) and their interactions, as well as random effects, including a random intercept and random slopes for day, course and measurement point.

Table 1
Descriptions of fixed and random effects ("*" implies both main effects and interactions), number of parameters (Num. Par.), Akaike information criterion (AIC) and Akaike weights (*w*(AIC)) for all included cumulative link mixed models. Akaike weights for the Models with a number 0 refer to this set of three models, for the four other models Akaike weights refer to their set of four models.

Model	Fixed effects	Random effects	Threshold constraints	Num. Par.	AIC	*w* (AIC)
$Model0a_{flexible}$	I + D * C * M	I + D + C + M	None (flexible)	30	6573.6	0.366
$Model0b_{equidistant}$	I + D * C * M	I + D + C + M	Equidistant	19	6651.1	0.000
$Model0c_{symmetric}$	I + D * C * M	I + D + C + M	Symmetric	24	6572.5	0.634
$Model1_{FixedNone}$	I	I	Symmetric	17	7241.1	0.000
$Model2_{FixedMain}$	I + D + C + M	I + D + C + M	Symmetric	20	6739.6	0.000
$Model3_{Full}$	I + D * C * M	I + D + C + M	Symmetric	24	6572.5	0.999
$Model4_{RandomIntercept}$	I + D * C * M	I	Symmetric	15	7568.6	0.000

I = intercept, D = day, C = course, M = measurement point.

$Model0a_{flexible}$ did not constrain the thresholds, thus the difference in the latent variable necessary to reach the next higher category could vary between each FMS category pair. In $Model0b_{equidistant}$, the increase necessary to be sorted into successively higher categories was the same for all category transitions, after a value of 1 had been reached. The last model, $Model0c_{symmetric}$, used symmetrical threshold constraints. This means for instance, that the increase necessary to progress one step up from a low FMS category is predicted to be the same as the necessary step increase when starting from a high FMS category.

A case for the analysis of FMS scores as an ordinal scale could be made if a cumulative link mixed model using equidistant thresholds was less parsimonious than otherwise identical models utilizing either symmetric (hypothesis H1a: AIC($Model0c_{symmetric}$) < AIC($Model0b_{equidistant}$)) or flexible thresholds (hypothesis H1b: AIC($Model0a_{flexible}$) < AIC($Model0b_{equidistant}$)). Based on the ideas expressed in the literature (Grimby et al., 2012), we also hypothesized that a model using symmetrical thresholds is preferable to a model that does not constrain its threshold estimates (hypothesis H1c: AIC($Model0c_{symmetric}$) < AIC($Model0a_{flexible}$)). In the resulting model comparisons, the related Akaike weights refer to this set of three models only.

The study's second aim was to analyze the development of VIMS symptoms over time. Concerning this aim, four mixed models, all using symmetric thresholds, were compared. Related Akaike weights refer to this set of four models.

$Model1_{FixedNone}$ did not include any fixed terms referring to day, course or measurement point. Compared with the full model's formula (Fig. 2), this model only included β_I as fixed and only $u_{I|i}$ as random effects. This model did therefore not predict any increases in FMS over time.

A second model ($Model2_{FixedMain}$) included fixed effect terms for day, course and measurement point and random effects that predicted normal distributed individual differences in these increases. Compared with the full model (Fig. 2), only the interaction terms were not included. A comparison between $Model1_{FixedNone}$ and $Model2_{FixedMain}$ examined the progression of FMS scores over time, but with always the same increases within and between courses (H2).

Here we expected that the incorporation of driving time in general would lead to preferable models (hypothesis H2: AIC($Model2_{FixedMain}$) < AIC($Model1_{FixedNone}$)) and in particular that FMS scores would increase over time within one drive (hypothesis H2a: fixed effect: measurement point), increase across drives on the same day (hypothesis H2b: fixed effect: course), and decrease on a subsequent day (hypothesis H2c: fixed effect: day).

Interaction terms were added in the full model ($Model3_{Full}$; Fig. 2). The inclusion of interaction terms in the model was predicted to lead to a better balance between model complexity and data fit (hypothesis H3: AIC($Model3_{Full}$) < AIC($Model2_{FixedMain}$)). Specifically, the increase of symptoms across the duration of one drive was thought to differ on successive drives on the same day (hypothesis H3a: interaction: Course ∗ Measurement point) and to be shallower on a subsequent day (hypothesis H3b: interaction: Day ∗ Measurement point). The increase of symptoms over successive drives observed on the first day was conjectured to be diminished on a subsequent day (hypothesis H3c: interaction: Day ∗ Course).

Any effects of day, course, and measurement were thought to persist when possible interactions are taken into account (hypothesis H3d), e.g., an absolute decrease in the level of symptoms between days would remain, beyond FMS scores just developing more slowly on the second day.

Lastly, $Model4_{RandomIntercept}$ included the full model's fixed effects but only random intercepts for each subject, dropping all random slopes. This means that only one term, $u_{I|i}$, was included as a random effect (compare Fig. 2), and thus each participant was predicted to show the same symptom progression over time, but at different general symptom levels. Random slopes, by contrast, predict a normal distribution of possible symptom progressions over time, with each participant's individual progression having been drawn from this distribution.

In comparing models with and without random slopes, the possible relevance of these individual developments was assessed. Participants were not only expected to display generally higher or lower symptom levels, but also different symptom developments over time (hypothesis H4: AIC($Model3_{Full}$) < AIC($Model4_{RandomIntercept}$)).

Lastly, the study aimed to find road segments that were more likely than others to result in increased FMS scores. For this purpose, the relative frequencies that an increase in FMS score would occur between two FMS measurement points, independently of the size of this difference, was calculated for all successive measurement point pairs. Specifically, quick up-down motions (bumpy road) were expected to be associated with higher frequencies of symptom increase, as compared

to other stretches of the road (hypothesis H5). To test this, a Friedman rank sum test was calculated with post-hoc analysis using Conover's test for a two-way balanced complete block design using the PMCMR package (Pohlert, 2014) in R 3.2.1. The tests employed Holm–Bonferroni corrections to control the familywise error rate.

3. Results

Four participants (three female) did not complete the experiment after reaching the FMS cut-off score of 15. All drop-outs occurred within the first course of the first day, with the respective experiments ending after the 4th, 15th, 16th, and 17th measurement points.

3.1. Visually induced motion sickness symptoms

The last FMS scores of each course were consistently positively correlated with the subsequently measured SSQ total scores with an average Kendall's $\bar{\tau}_b$ of 0.58 ($\tau_{course1} \approx 0.50$; $\tau_{course2} \approx 0.65$; $\tau_{course3} \approx 0.60$; $\tau_{course4} \approx 0.60$; $\tau_{course5} \approx 0.56$; $\tau_{course6} \approx 0.58$).

The symptoms most often reported in the SSQ included headache, nausea and vertigo. They differed among participants with low FMS scores (third with lowest observed FMS maxima <3; nine participants), whose reports included headaches and dizziness, but almost never nausea, and participants with relatively high FMS scores (third with highest observed FMS maxima >6; nine participants), who reported nausea, sweating and stomach awareness. No cases of vomiting were induced, and measured blood pressure remained normal.

The distribution of FMS scores (Fig. 3) diverged strongly from the normal distribution according to visual assessment and a Shapiro-Wilk-test, $W = 0.718$, $p < 0.01$. The scores were zero-inflated with 51.1% of the symptom scores being given as 0. Since the median as a measure of central tendency often remained at 0, the third quartiles will be reported in the following text.

3.2. Intervals between FMS categories (H1)

A comparison of three cumulative link mixed models differing only in their method of threshold estimation can indicate whether the increase in the latent variable necessary to raise the FMS score by 1 is best conceptualized as equidistant, flexible, or structured in a symmetrical manner. An overview of these models, and the comparisons between them, can be found in Table 1.

The required increases in the latent variable to progress to the next higher FMS category are shown in Fig. 4 for all three models. Here, the structure suggested by the symmetric threshold model coincided largely with the estimates for flexible thresholds. They showed an initial steep increase in thresholds, and consequently a slow increase in FMS scores, until a score of 5 was reached. Following this, the symptom scores were predicted to progress quickly, due the smaller rises in thresholds, until a score of 9. Finally, in the symmetric thresholds model, a corresponding slow increase in FMS scores was observed for

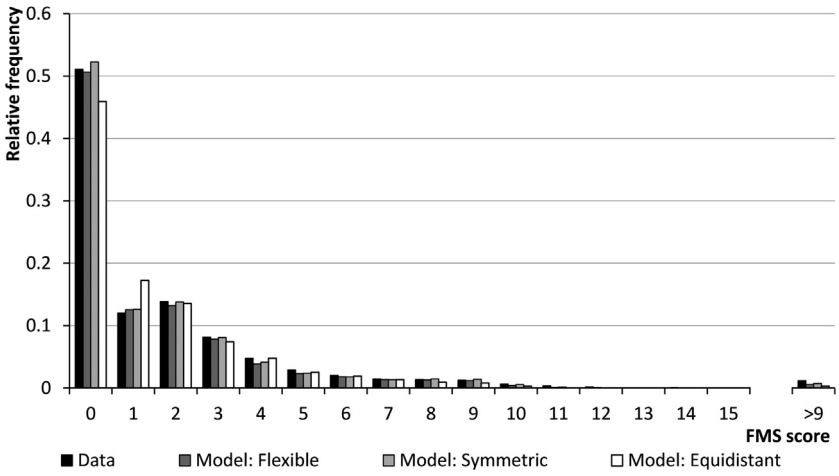

Fig. 3. Relative frequencies of reported visually induced motion sickness severity on the Fast Motion sickness Scale (FMS) for 28 participants, and relative frequencies of simulated expected values from three cumulative link mixed models. The three models all include the same fixed and random-effect structures but differ in their method for threshold estimation (Flexible: necessary increases in the latent variable to reach next higher FMS category differ for each threshold; Symmetric: necessary increases in the latent variable differ, but pairs of high and low thresholds with the same distance from a central category require equal rises; Equidistant: necessary increases in the latent variable are equal for all thresholds).

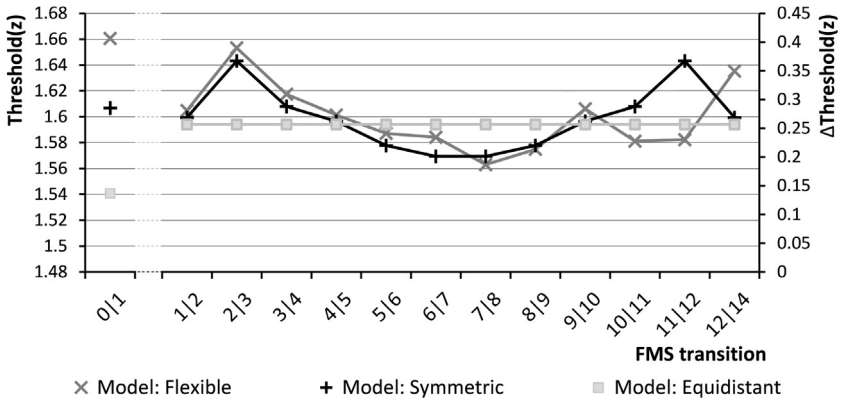

Fig. 4. Estimates from three cumulative link mixed models for the thresholds between two consecutive categories on the Fast Motion sickness Scale (FMS). All thresholds were standardized for easier comparability between models. Since the data was zero-inflated, the increase needed to pass the first threshold (0|1) is larger by almost an order of magnitude than for subsequent categories. Therefore, the left axis shows the increase in the latent variable needed to progress from the zero category to the one category, while the right axis indicates the additional increases needed for subsequent transitions between two FMS categories, which are displayed after the gap. The three models only differ in their method for threshold estimation (Flexible: necessary increases in the latent variable to reach next higher FMS category differ for each threshold; Symmetric: necessary increases in the latent variable differ, but pairs of high and low thresholds with the same distance from a central category require equal rises; Equidistant: necessary increases in the latent variable are equal for all thresholds). FMS scores of 13 were not observed.

transitions to values larger than 9. Here, flexible diverged from symmetric thresholds in unclear patterns, as increases were both larger and smaller than their counterparts. It should be noted that answers falling into FMS categories above 9 only constituted about 1.2% of all observations (see Fig. 3).

Likelihood ratio tests comparing the three models indicated that both the models utilizing symmetrically constrained (H1a; $\Delta AIC_{Symmetric\text{-}Equidistant} = -78.6$), $LR(5) = 102.080$, $p < 0.01$, or flexible thresholds (H1b; $\Delta AIC_{Flexible\text{-}Equidistant} = -77.5$), $LR(11) = 89.405$, $p < 0.01$, were to be preferred in comparison to the model using equidistant thresholds. Evidence ratios mark them as $2.03 * 10^{17}$ and $1.16 * 10^{17}$ more likely to be the best model among the three than the model using equidistant thresholds. Compared with the model using symmetric thresholds, using flexible thresholds did not significantly improve the model to allow for the more complex thresholds structure (H1c; $\Delta AIC_{Symmetric\text{-}Flexible} = -1.1$), $LR(6) = 9.574$, $p = 0.144$, and in fact the model using symmetric thresholds was 1.73 times more likely to be the better model.

The differences between the models can be demonstrated using the relative frequencies of the three models' predictions contrasted with the observed data, as shown in Fig. 3. If equal distances between the FSM categories were assumed, then the model underestimated the frequency of reports with either no symptoms or larger symptom scores, whereas FMS scores of 1 were predicted more often than they were observed. The models using flexible or symmetrical thresholds were generally comparable with one another, though the flexible thresholds model showed less divergence from the data for low symptom scores but more for higher FMS scores.

3.3. Time course of visually induced motion sickness

3.3.1. Development of visually induced motion sickness over time – inclusion of fixed effects for day, course and measurement point (H2)

The comparison of a model that included fixed effects for course, day, and measurement point ($Model2_{FixedMain}$) with a model without these terms ($Model1_{FixedNone}$) shows whether VIMS symptoms remain constant over time or change within each drive and between drives. As is shown in Table 1, the inclusion of the fixed effects led to a significantly improved model ($\Delta AIC_{Model1-Model2} = 501.5$) compared to the model treating VIMS as unchanging over time, $LR(3) = 507.510$, $p < 0.001$. Although Akaike weights indicated that neither model had a high probability of being the best among all considered models, the evidence ratio suggested that $Model2_{FixedMain}$ was $7.931 * 10^{108}$ times more likely to be the preferable model.

As seen in Table 2 the parameter coefficient estimates for $Model2_{FixedMain}$ show significant effects for day, course and measurement point. In the data (Fig. 5), the negative estimate for day (H2a) reflected a decrease in the observed 3rd quartiles of the FMS scores from 3 on the first to 2 on the second day. Within each day of the experiment the model indicated an increase over subsequent drives (H2b), shown in the data e.g. in increases of 2 categories in the 3rd quartiles between the two courses on the first day.

In each course, the model suggests a further increase over measurement points (H2c). This corresponds to increases from the first to the last measurement point in the observed 3rd quartiles for every course, e.g. 4 categories on the first course of the first day. It should be noted that the last measurement point was only the highest 3rd quartile during the last course.

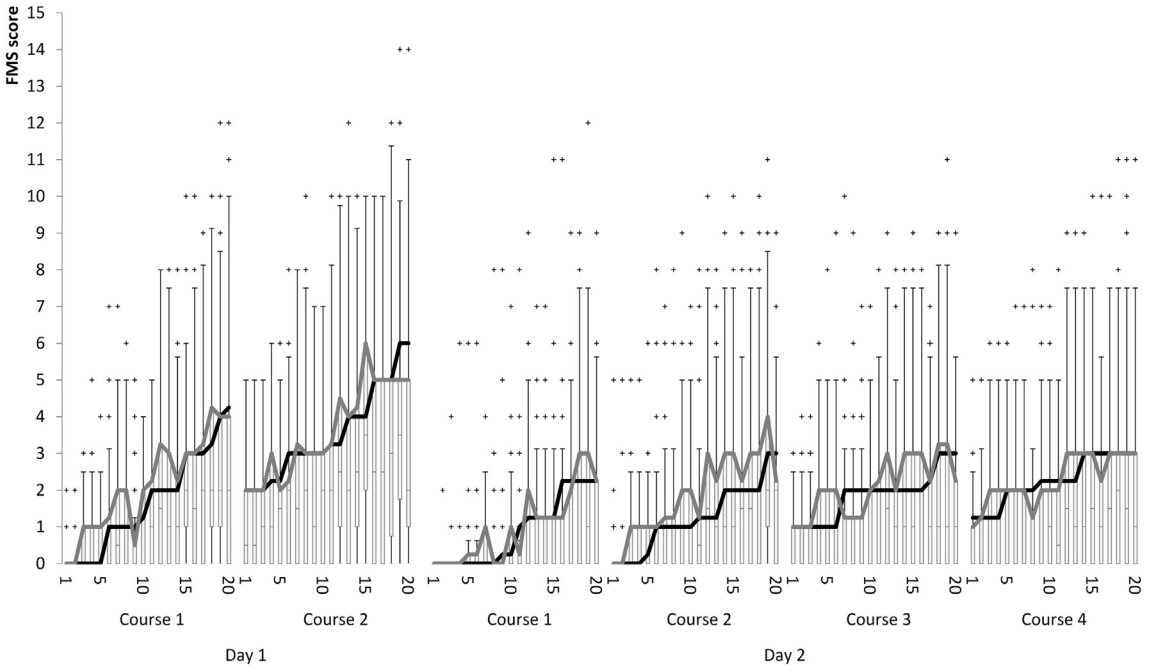

Fig. 5. Box-Whisker-Plot of the Fast Motion Sickness Scale (FMS) scores of 28 participants who completed all six courses. The whiskers extend 1.5 box lengths, unless this exceeds the highest or lowest observed values, respectively. All observations outside this range are plotted as plus signs. 3rd quartiles for the simulated expected FMS scores of $Model3_{Full}$ are displayed as black lines, while 3rd quartiles of the observed FMS scores are illustrated using grey lines.

3.3.2. Development of visually induced motion sickness over time – Additional inclusion of interaction terms (H3)

As the development of VIMS symptoms may not be uniform over the progression of the experiment's two days and six courses (H3), a comparison of the model including only fixed effects for day, course and measurement point ($Model2_{FixedMain}$) with a model that further considered possible interactions between these terms ($Model3_{Full}$) becomes relevant.

This further inclusion of the interaction terms between day, course and measurement point in the fixed effect structure led to a preferable model ($\Delta AIC_{Model2-Model3} = 167.1$) compared to the model without interaction terms. A likelihood ratio test showed this difference as highly significant, $LR(4) = 175.112$, $p < 0.001$. Akaike weights showed that $Model3_{Full}$ was likely the most parsimonious among the set of investigated models with the evidence ratio indicating this as being $1.92 * 10^{36}$ times more likely than in case of the model without interaction terms.

The estimates of $Model3_{Full}$ can be found in Table 2. While it still predicted a decrease in VIMS symptoms on the second day (Fig. 5), dropping from a predicted 3rd quartile value of 3 to a value of 2 on the second day, this was no longer explained by the fixed effect of day itself (H3d). Later courses on the same day were still associated with a rise in the FMS score (H3d), but this was shown to be stronger on the first day then on the second (H3c). In the data, the difference between the 3rd quartiles of the first two courses on the first day was 2 categories, while it was 1 on the second day.

Table 2
Estimates (β), standard errors (SE), 95% confidence intervals (95%-CI) and p-values for fixed effects concerned with the development of Fast Motion sickness Scale scores over time for two cumulative link mixed models ($N = 28$).

Model	Parameter	Fixed effect			
		β	SE	95%-CI	p
$Model2_{FixedMain}$	D	−2.801	0.670	[−4.117, −1.487]	< 0.001
	C	0.608	0.210	[0.197, 1.020]	0.004
	M	0.223	0.031	[0.162, 0.284]	< 0.001
$Model3_{Full}$	D	1.448	0.850	[−0.217, 3.113]	0.088
	C	5.114	0.676	[3.79, 6.439]	< 0.001
	M	0.494	0.093	[0.312, 0.675]	< 0.001
	DxC	−2.232	0.335	[−2.889, −1.575]	< 0.001
	DxM	−0.107	0.048	[−0.201, −0.014]	0.025
	CxM	−0.118	0.050	[−0.217, −0.02]	0.019
	DxCxM	0.045	0.026	[−0.007, 0.096]	0.087

D = day (2 levels), C = course (day 1: 2 levels; day 2: 4 levels), M = measurement point (levels 20 levels).

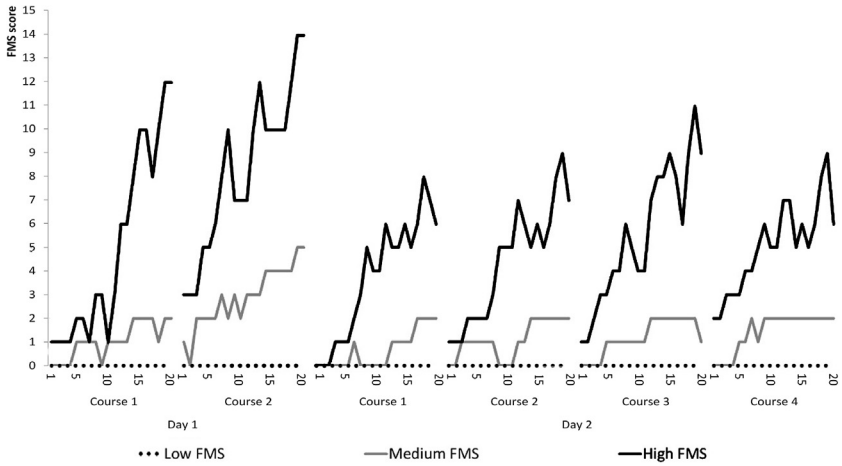

Fig. 6. FMS scores for three selected participants, one whose maximal FMS was among the smallest in the sample (dotted dark grey line), one whose maximal FMS equaled the medium (light grey line), and one whose maximal FMS was the highest in the sample (black line).

Within one drive, VIMS symptoms were still predicted to worsen (H3d), however, this was more pronounced on the first day (H3b) and in each day's earlier courses (H3a). In the data the difference between the 3rd quartiles of the first and last FMS measurements range from 4 categories during the first course of the first day, to 1.25 during the third course of the second day.

An effect of the interaction between day, course and measurement point could not be ascertained with a sufficient degree of probability (95%-CI [−0.007, 0.096]).

3.3.3. Individual differences in visually induced motion sickness susceptibility and progression (H4)

To illustrate the differences between participants, the FMS scores of three selected participants with low, medium or high reported symptom severity, are displayed in Fig. 6.

These differences can be explored via comparisons of mixed models with different random effects structures (H4). Here, a model with random intercepts and random slopes for each subject for day, course and measurement point ($Model3_{Full}$) was compared with a model with only random intercepts ($Model4_{RandomIntercept}$). The latter predicts the same development over time for each participant, though at a different symptom level. The model that included random slopes specified different progressions of VIMS symptoms over time for each individual participant.

The model including both random intercepts and random slopes showed a significantly better balance between model simplicity and data fit ($\Delta AIC_{Model4-Model3} = 996.1$) compared to $Model4_{RandomIntercept}$, $LR(9) = 1014.133$, $p < 0.001$. The evidence ratio indicated that the model with random slopes was $1.996 * 10^{216}$ times more likely to be the preferable model.

The random effects structure of $Model3_{Full}$ is reported in Table 3. Individuals differed in this model according to their intercept (range: −6.789, 11.471), the individual effects of day (range: −8.225, 4.594), course (range: −1.557, 1.862) and measurement point (range: −0.295, 0.400).

3.4. Effects of road segment (H5)

The roadway designed for this experiment included two stretches of road thought to induce VIMS, the bumpy road. These segments always fell between the same two FMS questions (Fig. 7). To answer whether these segments were especially likely to induce VIMS (H5), the relative frequencies of FMS increases were calculated from the six encounters (once each course) each participant had part of the course that was bounded by two FMS measurements.

Table 3
Random effect structure of $Model3_{Full}$ ($N = 28$; $N_{obs} = 3360$). This includes the standard deviations (*SD*) of each random effect term's estimated distribution and the estimated Pearson's correlations between the random effect terms.

Random effect	*SD*	Correlation		
		D	C	M
Intercept	4.069	−0.710	0.013	0.384
D	2.747		−0.216	−0.282
C	1.006			0.421
M	0.160			

D = day (2 levels), C = course (day 1: 2 levels; day 2: 4 levels), M = measurement point (levels 20 levels).

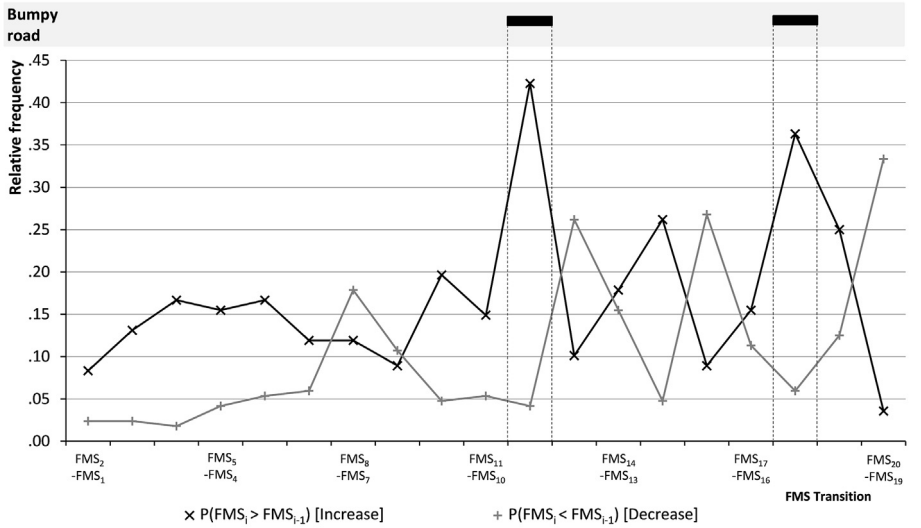

Fig. 7. Relative frequency of cases where two successive FMS scores were different, disregarding the size of the difference ($N = 28$). This includes both the relative frequency of later FMS questions indicating higher reported simulator sickness symptoms (black line) and of higher FMS scores being chosen for the earlier FMS question (grey line). Measurement intervals that included stretches of bumpy road are marked by black bars and dotted lines.

These segments of the track were more strongly associated with increases in the FMS scores, especially the first bumpy road segment (Fig. 7). In the two bumpy road segments, participants reported increased symptoms in on average 42.2% and 36.3% of their 6 encounters, making them here 10.143 and 6.100 times more likely to report an increase in the FMS score after this road segment than a decrease. A Friedman rank sum test on the relative frequencies of increases in the FMS score indicated that at least one of the segments differed from the others, χ^2 (18, $n = 28$) = 75.296, $p < 0.001$. Holm–Bonferroni-corrected pairwise tests for the first bumpy road segment showed that its frequency of a VIMS increase was significantly higher than 13 of the 17 road segments without a bumpy road. The second bumpy road was significantly more likely to produce an increase in FMS scores than 7 of the 17 non-bumpy road segments. The bumpy road segments were also identified as the most sickening elements in questionnaire answers concerning subjective perceptions of the course.

4. Discussion

The current study demonstrated the benefits of using the FMS as an on-line measure of VIMS severity during repeated drives over two days in a fixed-base driving simulator. No cases of extreme sickness symptoms such as vomiting were induced. We therefore conclude that the cut-off criterion of FMS score 15 was suitable for the prevention of frank sickness. We propose to use the FMS with a cut-off criterion of 15 as an effective and benign VIMS monitoring tool. Analysis of FMS scores using cumulative link mixed model has granted valuable insights into VIMS symptoms, their development over time, and their antecedents. Additionally, it was indeed beneficial to acknowledge the ordinal nature of the FMS data using non-equal thresholds in the cumulative link mixed models.

4.1. Optimal FMS score analysis strategy

The majority of participants experienced VIMS symptoms at some point during their simulator experience, with the majority of reported scores falling into lower FMS categories. For the most part, however, most subjects were symptom-free during long stretches of the driving sessions. In the current sample, the distribution deviated markedly from a normal distribution. Thus, the application of parametric tests would have been problematic. Based on our sample, we recommend to analyze the distribution of FMS data critically and to consider alternatives to parametric tests, such as nonparametric tests, or cumulative link mixed models (Christensen & Brockhoff, 2013).

For the sample at hand, the assumption that the distances between successive FMS categories were not equal, led to preferable models of FMS scores. These models estimated different necessary increases of the continuous latent variable to raise the FMS by one category, depending on the level of the pre-existing FMS score. Both a model considering these needed increases in the latent variable as potentially different for every transition between FMS categories (flexible thresholds; H1b), and a model postulating the increase to be symmetrically structured around a central category pair (symmetric thresholds; H1a), showed an improvement in their data fit that justified the higher model complexity. In comparing these two models, the model assuming symmetric thresholds was preferable (H1c). Symmetric thresholds have been hypothesized for similar scales in the literature (Grimby et al., 2012), and the frequent application of more common

parametric analysis strategies for this kind of data has been shown to potentially lead to incorrect inferences (Kahler et al., 2008). An analysis of FMS scores based on ordinal mixed models is not only an adequate response to the literature, but may, as was the case in our sample, benefit the analysis through better data fit and additional insights into the scale at hand.

4.2. Consequences for simulator sickness monitoring

For this sample, the symmetrical threshold structure could be subdivided into four parts: (1) the transition from FMS scores of 0 to 1, which needed higher increases in the latent variable compared to the rest of the scale, owing to the zero-inflated nature of the data; (2) the first four category transitions, i.e., a range from a FMS score of 1 until reaching a FMS score of 5, where the increase needed to advance to the next higher category was still relatively steep, corresponding to slow FMS score increases; (3) the middle four category transitions, i.e., starting from a FMS score of 5 until reaching a FMS score of 9, where FMS scores would be expected to rise quickly, as shown in the relatively low necessary increases in thresholds; and (4) the last observed category increases where the symptom scores should again rise relatively slowly, starting at a FMS score of 9 and ending at 14.

This analysis, however, only considered participants who completed the experiment, and in these participants, high FMS categories were seldom observed. This makes an interpretation of thresholds for high FMS scores difficult. Anecdotally, we can describe that two drop-outs in this study showed steep increases in FMS scores across the initial categories, with one aborting after about 5 minutes. A third participant showed a similarly steep increase after an initial plateau. When double-digit categories had already been reached, the experiment was also typically aborted shortly thereafter. Thus, participants who have to abort the experiment may well show different patterns, such as a fast development of symptoms across the entire FMS spectrum.

For less sensitive participants, the data suggest that once an FMS score of 5 is reached, further increases may come at an accelerated pace. When the FMS is used for VIMS monitoring, i.e., to regularly check the participant's status, the experimenter might want to consider whether an increased sampling frequency of FMS questions should be implemented above a score of 5. Since this can negatively impact the equality of observations, this should be well considered. Further studies are needed to clarify patterns in highly susceptible participants.

4.3. Development of visually induced motion sickness symptoms over time

For the current sample, the full model ($Model3_{Full}$) was the single most parsimonious model with the probability that it was the best among the tested models (Akaike weight) approaching one. This means that the temporal development of VIMS was best described by not only considering differences among the two experimental days, the six driven courses and the twenty measurement points per drive (H2), but that the interactions between these terms (H3) need to be included.

The reported FMS scores increase with time for each day of the experiment, both within one drive (H2a) and over the course of several drives (H2b). The increase over time within each course was, however, less pronounced during later drives compared to earlier ones (H3a). We interpret this interaction between course and measurement point as constituting habituation. To maximally facilitate desirable habituation, exposure times and pauses should be chosen strategically, with consideration for the time already driven during each day of simulator usage.

Participants were especially affected by VIMS during their first simulator usage (H2c). However, in this study, adaptation to VIMS seemed to be characterized by a different symptom development over the course of the first day compared to a week later, and less by symptoms developing in the same manner but at a generally lower level on the later date (H3d). Participant's adaptation to VIMS was shown in two factors. First, the increase of symptoms over the duration of each driven course was less pronounced a week after the first exposure (H3b). Second, the increase of VIMS symptoms across subsequent drives was lessened during the second experimental day (H3c). The markedly higher level of VIMS symptoms on the first day emphasizes that first time simulator users need special attention. Measures should be taken to monitor the symptoms of first time users and to prevent a steep FMS progression. For example, on first contact with the simulator the duration of rest periods between drives could be increased, scenarios could be shortened, or limited in their number, and provocative elements could be avoided.

It has been shown that there are substantial interindividual differences in participant's susceptibility to VIMS (Bock & Oman, 1982; Davis et al., 2015; Reason & Graybiel, 1970b). The current study suggests that this is not only constituted by generally higher or lower symptom levels, but by different symptom trajectories over time (H4). Using mixed models, the distributions of these differences can be taken into account in the analysis by utilizing both random intercepts, for generally higher or lower symptom scores, and subject-specific random slopes, to allow for differing symptom progressions over time.

4.4. Effect of driving scenario on visually induced motion sickness

One interesting finding is that the use of fine-grained FMS ratings made it possible to identify particularly sickening elements, i.e. the bumpy road. This is in line with studies showing the general influence of scenario design on simulator sickness (see Classen et al., 2011; Stoner et al., 2011). In the past, questionnaires have been used to evaluate the influence of scenario design encountered during driving sequences of e.g. 15 min. (Park et al., 2006). To the best of our knowledge, the current study is the first to identify a specific sickening element by means of an on-line motion sickness scale. This procedure

has some advantages. There is no possible recall bias because the symptom scores are obtained in close temporal proximity to the element under study. At the same time, the FMS is not very intrusive, allowing for its application during continuous driving. Depending on the scenario and the selection of test subjects, a quantitative analysis of the sickening effect seems possible. This study systematically looked at the impact of road elements producing increased vertical optical flow. However, the study of other factors, such as sharp curves, sharp braking, or roadside objects, could benefit from a similar analysis (Stoner et al., 2011).

4.5. Limitations

Our sample consisted only of young adults and may not generalize to older participants, who tend to be more sensitive to simulator sickness. This analysis further only considered participants who had completed the experiment. A generalization to extremely susceptible participants should not be taken for granted (see also Section 4.2). As these participants are rarely willing to come back for repeated testing, thorough adaptation in these participants may be very tedious to achieve (e.g. Dobie, May, Fischer, Elder, & Kubitz, 1987), but insights into how to best identify or accommodate this group in driving simulator experiments is of high practical relevance.

In the current study, the number of observations was too low to fit random slopes for interaction terms. Therefore, the individual differences could be captured more adequately in models based on more observations. Also, the additional symptom increases during the third and fourth courses on the second day are relatively small (Fig. 5), with further tendencies for the symptom increase during the drive to lessen with subsequent drives. This raises the question whether a plateau of sorts is reached after a certain number of drives, possibly on a lower symptom level for participants with simulator experience.

4.6. Conclusion

The use of the fine-grained FMS for on-line measurement of VIMS combined with cumulative link mixed models analysis has led to clear advantages for simulator sickness research, which is expected to extend to other motion sickness phenomena. Especially the choice to treat the FMS scores as ordinal data has proven beneficial in the given sample, and should be recommended whenever FMS scores clearly deviate from a normal distribution.

Using mixed models for ordinal data, we were able to reveal separate processes of short-term habituation and longer-term adaptation effects in the development of VIMS symptoms over time. Adaptation was evidenced by more gradual symptom increases on the second day, rather than merely parallel developments at lower symptom levels. The FMS could also be used to identify road elements that had a particularly strong impact on VIMS, opening the door for future research on simulator scenario design. Thus, the FMS in conjunction with mixed model analysis is a powerful and efficient research tool.

Acknowledgements

The study was funded by a grant from the Johannes Gutenberg-Universität Mainz. The Mahsan® Kombi/DOA6 for drug screening was provided by MAHSAN Diagnostika, Reinbek, Germany. We thank Stephanie Preuß and Sebastian Bernuth for their help with the experimental procedure. Parts of this work have been performed in the context of the medical dissertations of Patricia Hengstenberg and Ender Tutulmaz (in preparation).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience, 52*, 29–37.

Bagshaw, M., & Stott, J. R. (1985). The desensitisation of chronically motion sick aircrew in the Royal Air Force. *Aviation, Space, and Environmental Medicine, 56*, 1144–1151.

Bles, W., de Graaf, B., Bos, J. E., Groen, E., & Krol, J. R. (1997). A sustained hyper-g load as a tool to simulate space sickness. *Journal of Gravitational Physiology, 4*, P1–P4.

Bock, O. L., & Oman, C. M. (1982). Dynamics of subjective discomfort in motion sickness as measured with a magnitude estimation method. *Aviation, Space, and Environmental Medicine, 53*, 773–777.

Christensen, R.H.B. (2015). *Ordinal – Regression Models for Ordinal Data.* R package version 2015-6-28. <http://www.cran.r-project.org/package=ordinal/>.

Christensen, R. H. B., & Brockhoff, P. B. (2013). Analysis of sensory ratings data with cumulative link models. *Journal de la Societe Francaise de Statistique & Revue de Statistique Appliquee, 154*(3), 58–79.

Classen, S., Bewernitz, M., & Shechtman, O. (2011). Driving simulator sickness: An evidence-based review of the literature. *The American Journal of Occupational Therapy, 65*(2), 179–188.

Davis, S. L., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. In Y. Pisan, K. Nesbitt, & K. L. Blackmore (Eds.), *Proceedings of the 11th Australasian conference on interactive entertainment (IE 2015)* (pp. 3–14). Sydney: Australian Computer Society Inc. <http://crpit.com/confpapers/CRPITV167Davis.pdf>.

Diels, C., & Howarth, P. A. (2013). Frequency characteristics of visually induced motion sickness. *Human Factors, 55*, 595–604.

Dobie, T. G., May, J. G., Fischer, W. D., Elder, S. T., & Kubitz, K. A. (1987). A comparison of two methods of training resistance to visually-induced motion sickness. *Aviation, Space, and Environmental Medicine, 58*(9 Pt 2), A34–41.

Garrick-Bethell, I., Jarchow, T., Hecht, H., & Young, L. R. (2008). Vestibular adaptation to centrifugation does not transfer across planes of head rotation. *Journal of Vestibular Research, 18*, 25–37.

Golding, J. F., Mueller, A. G., & Gresty, M. A. (2001). A motion sickness maximum around the 0.2 Hz frequency range of horizontal translational oscillation. *Aviation, Space, and Environmental Medicine, 72*, 188–192.

Golding, J. F., & Stott, J. R. (1995). Effect of sickness severity on habituation to repeated motion challenges in aircrew referred for airsickness treatment. *Aviation, Space, and Environmental Medicine, 66*, 625–630.
Grimby, G., Tennant, A., & Tesio, L. (2012). The use of raw scores from ordinal scales: Time to end the malpractice. *Journal of Rehabilitation Medicine, 44*, 97–98.
Howarth, P. A., & Hodder, S. G. (2008). Characteristics of habituation to motion in a virtual environment. *Displays, 29*, 117–123.
Kahler, E., Rogausch, A., Brunner, E., & Himmel, W. (2008). A parametric analysis of ordinal quality-of-life data can lead to erroneous results. *Journal of Clinical Epidemiology, 61*, 475–480.
Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology, 3*(3), 203–220.
Kennedy, R. S., Stanney, K., & Dunlap, W. (2000). Duration and exposure to virtual environments: Sickness curves during and across sessions. *Presence: Teleoperators & Virtual Environments, 9*(5), 463–472.
Keshavarz, B., & Hecht, H. (2011). Validating an efficient method to quantify motion sickness. *Human Factors: The Journal of the Human Factors and Ergonomic Society, 53*(4), 415–426.
Keshavarz, B., Hecht, H., & Lawson, B. D. (2015). Visually-induced motion sickness: Causes, characteristics, and countermeasures. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments. Design, implementation, and applications* (2nd ed., pp. 647–698). New York: CRC Press Taylor.
Kolasinski, E.M. (1995). *Simulator sickness in virtual environments*. Technical report 1027. Alexandria: U.S. Army Research Institute for the Behavioral and Social Sciences. <www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA295861>.
Lawson, B. D. (2015). Motion sickness symptomatology and origins. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments. Design, implementation, and applications* (2nd ed., pp. 531–600). New York: CRC Press Taylor.
Lawther, A., & Griffin, M. J. (1986). The motion of a ship at sea and the consequent motion sickness amongst passengers. *Ergonomics, 29*, 535–552.
Mackrous, I., Lavallière, M., & Teasdale, N. (2014). Adaptation to simulator sickness in older drivers following multiple sessions in a driving simulator. *Gerontechnology, 12*, 101–111.
McCauley, M. E., Royal, J. W., Wylie, C. D., O'Hanlon, J. F., & Mackie, R. R. (1976). *Motion sickness incidence: Exploratory studies of habituation, pitch, roll and the refinement of a mathematical model* Technical report 1733-2. Goleta, CA: Human Factors Research Division, Canyon Research Group Inc.
Mourant, R. R., Rengarajan, P., Cox, D., Lin, Y., & Jaeger, B. K. (2007). The effect of driving environments on simulator sickness. In *Proceedings of the human factors and ergonomics society 51st annual meeting* (pp. 1232–1236). Santa Monica, CA: Human Factors and Ergonomics Society.
Muttray, A., Breitinger, A., Goetze, E., Schnupp, T., Geissler, B., Kaufmann, T., & Letzel, S. (2013). Further development of a commercial driving simulation for research in occupational medicine. *International Journal of Occupational Medicine and Environmental Health, 26*(6), 1–17.
Park, G. R., Allen, W., Fiorentino, D., Rosenthal, T. J., & Cook, M. L. (2006). Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study. In *Proceedings of the human factors and ergonomics society 50th annual meeting* (pp. 2702–2706). Santa Monica, CA: Human Factors and Ergonomics Society.
Pohlert, T. (2014). *The pairwise multiple comparison of mean ranks package (PMCMR).* R package version 2016-01-06. <http://CRAN.R-project.org/package=PMCMR>.
Reason, J. T., & Graybiel, A. (1970a). Progressive adaptation to Coriolis accelerations associated with 1-rpm increments in the velocity of the slow rotation room. *Aerospace Medicine, 41*, 73–79.
Reason, J. T., & Graybiel, A. (1970b). Changes in subjective estimates of well-being during the onset and remission of motion sickness symptomatology in the slow rotation room. *Aerospace Medicine, 41*, 166–171.
Stoner, H. A., Fisher, D. L., & Mollenhauer, M. A. (2011). Simulator and scenario factors influencing simulator sickness. In D. Fisher, M. Rizzo, J. Caird, & J. Lee (Eds.), *Handbook of driving simulation for engineering, medicine and psychology*. Boca Raton, FL: CRC Press.
Tyler, D. B., & Bard, P. (1949). Motion sickness. *Physiological Reviews, 29*, 311–369.
Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*, 192–196.
Watson, G. S. (2000). A synthesis of simulator sickness studies conducted in a high-fidelity driving simulator. In *Proceedings of the driving simulation conference (DSC)* (pp. 69–78). Paris: Driving Simulation Association.
West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: A practical guide using statistical software*. New York: Chapman & Hall/CRC.