

## Abschlussberichts zum Vorhaben

# Achtung Blaulicht – Evaluation des Trainings „Verkehrssicherheit bei Einsatzfahrten“ des DVR und der DGUV (FP-0366)

Laufzeit

01.06.2014 – 30.11.2017

Bericht vom 05.03.18

Autoren

Maria Prohn

Dennis Nowak

Britta Herbig

In diesem Bericht wird versucht, soweit wie möglich geschlechtsneutrale Wortformen zu verwenden, die sowohl Frauen als auch Männer bezeichnen sollen. Wenn dies nicht möglich ist, wird in der Regel zugunsten der besseren Lesbarkeit nur eine Geschlechtsform benutzt, die jeweils nicht verwendete Form ist jedoch implizit miteinbezogen.

## **Projektteam**

Ines Englmann, Studienassistentz

PD Dr. phil. Dr. habil. med. Britta Herbig, Studienleitung

Prof. Dr. med. Dennis Nowak, Studienleitung

Maria Prohn, Dipl.-Psych., Projektmitarbeiterin

Institut und Poliklinik für Arbeits-, Sozial- und Umweltmedizin, Klinikum der Universität München, Ziemssenstr. 1, D-80336 München

## **Kooperationspartner**

Deutscher Verkehrssicherheitsrat e.V., Auguststr. 29, D-53229 Bonn

Institut für Arbeit und Gesundheit der Deutschen gesetzlichen Unfallversicherung e.V., Königsbrücker Landstraße 2, D-01109 Dresden

## **Projektförderer**

Deutsche Gesetzliche Unfallversicherung e. V., Alte Heerstraße 111, 53757 St. Augustin

Wir möchten uns bei allen teilnehmenden Organisationen und ihren Ansprechpartnern für die Unterstützung der Studie herzlich bedanken. Unser besonderer Dank gilt allen Rettungsdienstmitarbeitern, die bereit waren, trotz ihres anstrengenden Arbeitsalltags an unserer aufwändigen Studie aktiv teilzunehmen.

## Inhaltsverzeichnis

Kurzfassung.....	1
Summary .....	3
1. Problemstellung.....	5
2. Forschungszweck/-ziel .....	11
3. Methodik.....	15
3.1 Studiendesign.....	15
3.2 Datenerhebungsmethoden .....	16
3.2.1 Fragebogendaten .....	16
3.2.1.1 Reaktionsebene.....	17
3.2.1.2 Lernebene .....	17
3.2.1.3 Verhaltensebene .....	20
3.2.1.4 Resultateebene .....	20
3.2.1.5 Kontrollvariablen.....	21
3.2.2 Wissenstest: Entwicklung und Auswertung.....	23
3.2.3 Fahrprofile und Videodaten.....	27
3.2.4 Schichtbegleitmessungen: Einsatzdaten, EKGs und Beanspruchung während der Schicht .....	29
3.3 Analysemethoden.....	31
3.4 Stichprobe und Zellbesetzungen.....	32
3.4.1 Akquise.....	32
3.4.2 Stichprobe und Dropout-Analyse .....	33
3.4.3 Zellbesetzungen auf den Evaluationsebenen und zu den Messzeitpunkten.....	35
3.4.4 Beschreibung der Stichprobe.....	37
3.4.5 Verteilung der Kontrollvariablen in den Gruppen.....	39
3.5 Beschreibung der Ausgangswerte vor dem Training.....	45
4. Ergebnisse des Gesamtvorhabens .....	50
4.1 Beschreibung der Trainingszusammensetzungen .....	50
4.2 Hauptanalyse: Ergebnisse zu Effekten des Trainings und der Messungen .....	51
4.2.1 Reaktionen auf die durchgeführten Trainings.....	54
4.2.1.1 Haupteffekte in der Reaktionsebene direkt nach dem Training .....	54
4.2.1.2 Reaktionen auf das Training im langfristigen Verlauf .....	56
4.2.1.3 Freie Kommentare zum Training.....	57
4.2.2 Vergleich aller drei Studiengruppen zu den Zeitpunkten T1 und T2 .....	60
4.2.2.1 Haupteffekte in der Lernebene .....	61
4.2.2.2 Haupteffekte in der Verhaltensebene.....	72
4.2.2.3 Haupteffekte in der Resultateebene.....	80
4.2.3 Prüfung der Effekte der Messungen .....	89
4.2.4 Prüfung der Voraussetzung zur Zusammenführung der Messzeitpunkte .....	92
4.3 Weitere Ergebnisse zur Lernebene.....	95
4.3.1 Lernen durch das Training in der Kontrollgruppe .....	96
4.3.2 Lernen im langfristigen Verlauf .....	98
4.3.3 Lernen durch das Training in der ungepaarten Stichprobe.....	99
4.4 Weitere Ergebnisse zur Verhaltensebene.....	102
4.4.1 Verhaltensänderung durch das Training in der Kontrollgruppe .....	102
4.4.2 Verhaltensänderung im langfristigen Verlauf .....	105
4.4.3 Verhaltensänderung durch das Training in der ungepaarten Stichprobe.....	106
4.5 Weitere Ergebnisse zur Resultateebene.....	110
4.5.1 Veränderung der Resultate durch das Training in der Kontrollgruppe.....	110

4.5.2	Veränderung der Resultate im langfristigen Verlauf .....	114
4.5.3	Veränderung der Resultate durch das Training in der ungepaarten Stichprobe .....	117
5.	Moderationseffekte von Erfahrung .....	120
5.1	Einfluss der Erfahrung auf kurzfristige Trainingseffekte .....	120
5.2	Einfluss der Erfahrung auf langfristige Trainingseffekte .....	121
6.	Erfolgsfaktoren .....	124
6.1	Bildung von Indices zum allgemeinen und ebenen-spezifischen individuellen Trainingserfolg .....	124
6.2	Anzahl erfolgreicher Teilnehmer .....	125
6.3	Prädiktion von Erfolg durch individuelle Voraussetzungen .....	128
6.4	Prädiktion von Erfolg durch organisationsbezogene Voraussetzungen .....	138
6.5	Prädiktion von Erfolg durch gruppenbezogene Voraussetzungen .....	146
7.	Einflussfaktoren auf Sondersignalfahrten im Rettungsdienst .....	157
7.1	Einsatz-Hinfahrten im Vergleich zu Patiententransporten .....	157
7.2	Einfluss von Schichtvariablen auf Sondersignalfahrten .....	158
7.3	Videoanalysen .....	161
7.3.1	Entwicklung des Analyseverfahrens zur Videosichtung .....	161
7.3.2	Potenziell kritische Situationen bei Sondersignalfahrten .....	163
7.3.3	Kategorien potenziell kritischer Situationen .....	165
7.4	Einflussfaktoren auf Beanspruchung im Rettungsdienst .....	171
8.	Zusammenfassung der Ergebnisse .....	173
9.	Diskussion und praktische Implikationen .....	176
10.	Kommunikation und Umsetzung der Ergebnisse .....	179
Literaturverzeichnis .....		181
Abbildungsverzeichnis .....		186
Tabellenverzeichnis .....		188
Anhang/Anhänge .....		190
Anhang A .....		191
Anhang B .....		194

## Kurzfassung

*Hintergrund:* Das Risiko eines Unfalles während Sondersignalfahrten ist im Vergleich zu Normalfahrten deutlich erhöht (17-fach für Sachschäden, 8-fach für Schwerverletzte, 4-fach für Todesfolgen) und ca. alle 272.000 Einsatzfahrten kommt es zu einem tödlichen Unfall. Kontinuierlich steigende Einsatzzahlen verschärfen dieses Problem und ein hoher Anteil der Unfälle bei Sondersignalfahrten geht auf die Fahrer der Einsatzfahrzeuge zurück. In einem Berufsfeld mit dem Hauptziel, Menschen zu retten, ist die zusätzliche Gefährdung im Straßenverkehr ein großes Problem und in doppelter Hinsicht tragisch: Neben der Gefahr für die am Unfall Beteiligten kann auch am eigentlichen Einsatzort erst später Hilfe erfolgen. Zur Erhöhung der Verkehrssicherheit von Einsatzfahrten erscheint es demzufolge sinnvoll, bei den Einsatzfahrern anzusetzen. Derzeit gibt es in Deutschland keine einheitliche Schulung oder Ausbildungsvorschrift für Fahrzeugführer von Sondersignalfahrzeugen. Der Verkehrssicherheitsrat und die Deutsche Gesetzliche Unfallversicherung haben vor einigen Jahren ein eintägiges simulatorbasiertes Verkehrssicherheitstraining speziell für Einsatzfahrten und -fahrer entwickelt. Theoretische Inhalte dieses Trainings sind neben den rechtlichen Grundlagen die Fahrphysik, Wahrnehmung, Einstellung und Gefahrenpotenziale besonders bei Einsatzfahrten. Mit Fahrten im Simulator erleben die Trainingsteilnehmer Konsequenzen dieser Hintergründe in Fahrsituationen, die beim Einsatz ihrer Fahrzeuge auftreten können, geben sich gegenseitig Feedback und arbeiten unter Anleitung speziell ausgebildeter Trainer gemeinsam Handlungsalternativen aus, die im realen Straßenverkehr angewendet werden können.

*Zielsetzung:* Ziel der vorgelegten Studie war es, herauszufinden, wie sich das oben genannte Training auf die Verkehrssicherheit von Sondersignalfahrten im Rettungsdienst auswirkt. Dazu wurde eine multimethodale Evaluation für verschiedene Aspekte der vier Ebenen Reaktion, Lernen, Verhalten und Resultate durchgeführt. Der Einfluss weiterer Faktoren, wie etwa Erfahrung der Fahrer oder Zusammensetzung der Trainingsgruppen, auf die Wirksamkeit des Trainings wurden überprüft und Verbesserungsmöglichkeiten und notwendige Rahmenbedingungen für den Trainingserfolg abgeleitet.

*Design und Methoden:* Um die für die Evaluierung notwendige eindeutige Rückführung der Effekte auf das Training zu ermöglichen, wurde als erweitertes Interventions-/Kontrollgruppendesign gewählt, bei dem die Kontrollgruppe als Wartegruppe ausgelegt war und eine zusätzliche Gruppe zur Untersuchung reiner Messeffekte herangezogen wurde. Mit einer kontrollierten Prä-Post-Messung (T1-T2) zur Wirksamkeitsprüfung, einer Wartegruppenerhebung (T2-T3) zur Erfassung der Replizierbarkeit und einer weiteren längerfristigen Messung (T4) zur Erhebung der Nachhaltigkeit von Effekten, wurde ein aufwändiges quasiexperimentelles Untersuchungsdesign mit bis zu vier Messzeitpunkten realisiert. Zur Sicherstellung der Generalisierbarkeit wurde das Training in Sachsen sowie im südlichen Bayern jeweils im ländlichen und städtischen Raum durchgeführt. Alle Trainings wurden von demselben Trainer durchgeführt. Teilnehmen konnten alle Rettungsdienstmitarbeiter, die mindestens 40 Stunden im Monat Rettungstransportwägen, Notarzteinsatzfahrzeuge oder Notarztwägen fahren und deren Rettungsdienstorganisation sich zur Teilnahme an der Studie bereit erklärt hatte. Alle Probanden wurden möglichst gleichmäßig auf die Untersuchungsgruppen aufgeteilt. Methodisch wurden die Evaluationsebenen wie folgt erfasst: Die Reaktionsebene, die die Akzeptanz und Zufriedenheit mit dem Training beinhaltet, wurde mittels Befragungen direkt nach dem Training und zu unterschiedlichen späteren Zeitpunkten erhoben. Die Lernebene mit den Aspekten der Einstellungsänderung und des Wissenszuwachses wurde mit Fragebögen zu den Themen Risikobereitschaft und Verkehrssicherheit sowie einem wiederholten Wissenstest erfasst. Für die Verhaltensebene wurden neben weiteren Fragebögen reale Einsatzfahrten mithilfe von Fahrprofilen (u.a. Quer- und Längsbeschleunigung, Geschwindigkeiten) und Videoaufzeichnungen untersucht. Auf der Resultatebene wurden die Belastung und Beanspruchung während des realen Einsatzgeschehens untersucht sowie die Dauer der Einsatzfahrten als wirtschaftliche Kenngröße. Die Erfassung der Bean-

spruchung erfolgte sowohl mit (schichtbegleitenden) Fragebögen zu kurzfristigen und langfristigen Beanspruchungsfolgen als auch mittels der Messung physiologischer Daten (EKGs) während der Einsätze.

*Ergebnisse:* Insgesamt 26 Organisationen nahmen an der Studie teil, für bis zu 182 Rettungsdienstmitarbeiter liegen Daten auf den unterschiedlichen Ebenen und zu den verschiedenen Messzeitpunkten vor. Insgesamt konnten 147 Personen vollständig zu den Zeitpunkten T1 und T2 erfasst werden, 30 davon auch in den Follow-up-Messungen. Die Teilnehmer waren zum ersten Messzeitpunkt durchschnittlich  $33.1 \pm 9.4$  Jahre alt, 21.9% waren weiblich. Der weitaus größte Teil (71.6%) war als Rettungsassistent ausgebildet. Im Erhebungszeitraum von Oktober 2014 bis Juni 2017 wurden in insgesamt 1468 Rettungsdienstschichten Messungen durchgeführt. Auf der Reaktionsebene zeigte sich, dass die Teilnehmer das Training direkt nach der Durchführung sehr positiv und langfristig noch recht gut bewerteten. Auf der Lernebene konnte durch das Training ein signifikanter Wissenszuwachs erzielt werden und die Risikowahrnehmung der Teilnehmer v.a. bei Normalfahrten wurde durch das Training erhöht. Dieser Einstellungseffekt war in der Wartegruppe replizierbar, im langfristigen Verlauf jedoch nicht nachhaltig. Auf der Verhaltensebene fanden sich in den Fahrprofilen Verringerungen der mittleren und maximalen Geschwindigkeiten bei Sondersignalfahrten durch das Training. Diese Veränderungen der Geschwindigkeit erwiesen sich sowohl als replizierbar wie auch als nachhaltig. Längerfristig zeigte sich auch eine Verringerung der Längsbeschleunigung. Belastbare Effekte auf der Resultateebene fanden sich nicht, allerdings ist bemerkenswert, dass sich trotz veränderter Geschwindigkeiten die durchschnittliche Dauer der Einsatzfahrten nicht verlängerte. Alle weiteren gemessenen Parameter der Ebenen Lernen, Verhalten und Resultate zeigten keinen Trainingseffekt. Auswertungen zu Kontroll- und weiteren potenziellen Einflussfaktoren zeigten häufig differenzielle, aber nicht uniforme, Einflüsse des Alters und Geschlechts der Teilnehmer auf den Trainingserfolg. Ein klarer Einfluss der Messung wurde ebenso wenig gefunden wie ein Einfluss der Erfahrung der Teilnehmer - besonders die Fahrerfahrung spielte keine Rolle. Hinsichtlich des Trainingserfolges finden sich nur wenige prädiktive Faktoren: Probanden mit tendenziell höherer Rettungsdienstenerfahrung und geringer Freude am Fahren mit Sondersignal sowie eher heterogen zusammengesetzte Trainingsgruppen profitierten eher vom Training. In Zusatzauswertungen der Fahrprofil- und Videodaten der Einsatzfahrten zeigten sich u.a. im Vergleich zu Fahrten zum Einsatzort deutlich geringere Quer- und Längsbeschleunigungen sowie geringere maximale Geschwindigkeiten, wenn Patienten mit Sondersignal transportiert wurden, wobei sich die mittlere Geschwindigkeit nicht unterschied. Die Videoanalysen zeigten, dass es alle 36.9 Sekunden zu einer potenziell kritischen Situation kam mit einer mittleren Dauer von 6 Sekunden. Die häufigsten potenziell kritischen Situationen waren dabei Vorfahrtsregelungen gefolgt von Überholvorgängen.

*Zusammenfassung und Diskussion:* Das Training zeigte insgesamt nur in wenigen erfassten Aspekten signifikante Effekte in die erwünschte Richtung und keiner der anderen Aspekte veränderte sich in eine für die Verkehrssicherheit kritischere Richtung. Neben positiven Reaktionen auf das Training, wurden 4 der 68 gemessenen Variablen der Ebenen Lernen, Verhalten und Resultate durch das Training verändert. Allerdings wurden dabei positive Entwicklungen in zwei sehr wichtigen Verhaltensaspekten, der mittleren und der maximalen Geschwindigkeit, gefunden. Da (überhöhte) Geschwindigkeit zu einem der größten Einflussfaktoren im Unfallgeschehen zählt, sollte mit einer Reduzierung eine Erhöhung der Verkehrssicherheit bei Einsatzfahrten einhergehen. Fasst man alle Ergebnisse des Projekts zusammen, muss abschließend konstatiert werden, dass das Training belastbare (replizierbare und nachhaltige) und positive, wenn auch kleine, Effekte erzielt und ggf. durch einige Maßnahmen (wie z.B. heterogen zusammengesetzte Trainingsgruppen) in seiner Wirksamkeit verbessert werden kann. In einem so wichtigen Gebiet wie der Verkehrssicherheit bei Einsatzfahrten, in dem neben dem Fahrer nicht nur andere Verkehrsteilnehmer, sondern auch Hilfsbedürftige von einer sicheren Fahrweise abhängen, sollten auch solche Effekte als Erfolg verbucht werden.

## Summary

**Background:** The risk for an accident during blue light driving compared to normal driving is significantly higher (17x higher risk for material damage, 8x higher risk for seriously injured, and 4x higher risk for fatal accidents). About every 272.000<sup>th</sup> emergency drive a fatal accident happens. An increasing number of operations exacerbates the problem and a high number of accidents traces back to the driver. In an occupational area with the main goal of saving lives this additional traffic risk is a big problem and twofold tragic: Besides the danger for all involved in an accident, help arrives delayed at the primary emergency site. To increase road safety for emergency drives it seems therefore wise to start with the drivers. At the moment there is no uniform and standardized training course for drivers in Germany; however, the German Traffic Safety Council and the German Social Accident Insurance developed a one-day simulator-based training especially for blue light driving. Theoretical contents of this training are legal basics, driving physics, perception, attitudes and risk potentials during blue light drives. Participants experience the consequences of these topics during simulated drives in situations that may happen during real emergency operations. They give feedback to one another and, under the guidance of specifically qualified trainers, work out alternatives of taking action in real-life driving situations.

**Aims:** Goal of our project was to analyze how this training affects road safety during blue light driving in emergency medical services. Therefore, a multi-method evaluation of this training on the levels of reaction, learning, behavior, and results was conducted and the influence of further factors, like e.g. experience of the drivers or composition of the training groups, on the effectiveness of the training was tested. Options for potential improvements and necessary preconditions should also be deduced.

**Design and methods:** To make sure that effects can be unambiguously attributed to the training an expanded intervention-control-group design was used. The control group was realized as waiting group and a further group was used to capture pure effects of measurement. Thus, an elaborate quasi-experimental design with up to four measurement times was used including a controlled pre-post-measurement (T1-T2) for the test of effectiveness, a waiting group measurement (T2-T3) for a test of replicability, and a fourth long-term measurement (T4) for a test of sustainability of effects. To ensure generalizability trainings were conducted in Saxony and southern Bavaria, Germany, in both urban and rural areas. All trainings were done by the same trainer. All employees of emergency medical services could participate as long as they drove at least 40 hours a month emergency ambulances or emergency doctor's vehicles and as long as their organization was willing to participate in the study. All participants were allocated to the experimental groups as evenly as possible. Methods used to measure changes on the different levels of evaluation were as follows. On the level of reaction participants had to fill in a survey of acceptance and satisfaction with the training directly afterwards and at later measurement times. For the level of learning questionnaires on attitudes regarding road safety and risk propensity were used as well as a newly developed knowledge test. Besides further questionnaires the level of behavior was measured using aspects of driving behavior in real traffic via telemetric data tracking (e.g., speed, longitudinal and lateral accelerations) and video recording. Finally, on the level of results data on stress and strain during real shifts and emergency operations were collected by physiological data (ECGs) and (shift-related) questionnaires on short- and long-term strain reactions. As an economic key performance indicator the length of emergency drives was measured.

**Results:** Overall, 26 rescue service organizations participated in the study and data from up to 182 employees are available for the different evaluation levels and measurement times. 147 participants were measured completely at T1 and T2 of which 30 could also be included in the follow-up. At first measurements participants were 33.1±9.4 years on average. 21.9% were female and most (71.6%) were qualified as paramedics. In the data collection period between October 2014 and June 2017 a total of 1468 working shifts were recorded. On the level of reaction, participants evaluated the training very good directly after the training and quite good in the long run. A significant knowledge gain was achieved on the

level of learning and risk perception especially for normal drives was increased by the training. This effect on attitudes was replicated in the waiting group; however, it was not sustainable. On the level of behavior the training caused a decrease in mean and maximum speed during real emergency drives. These changes were replicable as well as sustainable in the long-term measurements. In the long run a decrease in longitudinal acceleration was also found. Sound proof for effects on the level of results could not be found. Strikingly, though, there was no difference in the mean duration of emergency drives despite the decreases in speed. For all other parameters that were measured on the levels of learning, behavior and results no significant effects of the training were found. Analyses regarding control and other potentially relevant aspects often showed differential but not uniform effects of age and sex on the success of the training. Moreover, clear-cut effects of the measuring situation or of the experience of participants, especially driving experience, were not found either. There were only few predictors for training success: Participants with (slightly) higher experience in emergency medical services and with less blue light driving pleasure as well as heterogeneously composed training groups profited more by the training. Additional analyses of the driving and video data showed, for example, that patient transports with blue light were done with significant lower maximum speed and lower longitudinal and lateral accelerations than blue light drives to an emergency site although mean speed did not differ. Video recordings showed that during emergency drives every 36.9 seconds a potentially critical situation with an average duration of 6 seconds happens. The most common potentially critical situations pertained to right-of-way laws and overtaking maneuvers.

*Summary and Discussion:* Overall, the evaluation showed only few significant desired training effects and at least none of the other aspects changed critically with regard to road safety. In addition to positive reactions to the training 4 out of 68 measured variables on the levels of learning, behavior and results changed due to the training. Indeed, positive developments were seen in two very important behavioral aspects – in mean and maximum speed. As (excessive) speed is one of the biggest factors in traffic accidents, a reduction should increase road safety during emergency drives. Taken all results into account, the training shows (replicable and sustainable) positive but small effects and might be improved by some actions (like, e.g. heterogeneous composed training groups). In an area as important as road safety during emergency drives that not only affects drivers and other traffic but also people in need even such few and small effects should be seen as an achievement.



## 1. Problemstellung

Seit Jahren nehmen die Einsatzzahlen im Rettungsdienst kontinuierlich zu, sowohl gesamt als auch bezogen auf die Einwohner. Bei der Überprüfung des Leistungsniveaus des öffentlichen Rettungswesens, die regelmäßig durch die Bundesanstalt für Straßenwesen (BASt) durchgeführt wird, zeigen sich die Steigerungen der Einsatzzahlen in Deutschland deutlich: Im Zeitraum 2012/2013 wurden inklusive Krankentransporten mittels Hochrechnungen zwölf Millionen Einsätze mit insgesamt 14,3 Millionen Einsatzfahrzeugen erfasst (Schmiedel, 2015). Das sind mehr als doppelt so viele Einsätze wie 1985 (5,97 Millionen) dokumentiert wurden. Mit Ausnahme der Krankentransporte werden dabei alle Einsatzarten (internistischer Notfall, Verkehrsunfall, Arbeitsunfall, sonstige Notfälle, sonstige Unfälle) in 75%-93% der Fälle mit Sondersignal durchgeführt. Im Vergleich zu 2008/2009 gab es 2012/2013 1,2 Millionen mehr Einsatzfahrten mit Sonderrechten, was einen Anstieg von 18% und insgesamt 8,4 Millionen Einsatzfahrten bedeutet (Schmiedel, 2015). Auch der Rettungsdienstbericht Bayern (INM, 2016) verdeutlicht einen Anstieg der Einsatzzahlen. So sind beispielsweise die absoluten Notfallereignisse in Bayern in den letzten 10 Jahren um über 50% angestiegen (INM, 2016). Bei 44% der Einsätze wird zusätzlich zum Rettungsdienst auch ein Notarzt benötigt (INM, 2016). Dies erfordert demzufolge in vielen Fällen mehr als ein anführendes Fahrzeug, denn fast alle Notarzteinsätze werden mittels Rendezvous-System, d.h. Zusammentreffen von Rettungswagen (RTW) und Notarzt erst am Einsatzort, durchgeführt (Schmiedel, 2015).

Unfälle mit Einsatzfahrzeugen werden nicht systematisch erfasst und veröffentlicht, daher gibt es keine belastbaren, aktuellen Zahlen, wie häufig diese auftreten. Einige Quellen gehen von deutlich erhöhten Unfallrisiken für Einsatzfahrzeuge aus: 4-mal häufiger für Todesfolge, 8-mal häufiger für Schwerverletzte und 17-mal häufiger für Sachschäden über 1500 Euro (Bockting, 2007; Müller, 2003; Pieper-Nagel & Wiegand, 2011; Unterkofler & Schmiedel, 1994). All diese Zahlen gehen auf die gleiche Quelle mit Zahlen aus dem Jahr 1991 (Unterkofler & Schmiedel, 1994) zurück und liegen damit inzwischen recht weit in der Vergangenheit. Ob sich die Wahrscheinlichkeit, unter Nutzung von Sondersignalen in einen Unfall verwickelt zu werden, seitdem verändert hat, ist nicht bekannt. In der allgemeinen deutschen Unfallstatistik steigen zumindest die Unfallzahlen kontinuierlich an bei immerhin insgesamt sinkenden Unfällen mit Todesfolgen.

In den Medien sind Unfälle mit Einsatzfahrzeugen jedoch häufig vertreten. Der Artikel „Hessen: Jeden Tag 5 Blaulicht-Unfälle!“ (2015) beschreibt, dass im Jahr 2015 beispielweise 1600 Unfälle mit Rettungsfahrzeugen registriert wurden, davon 200 mit Sondersignal (Frankfurter Neue Presse, 2015) und Hamburg wies 2010 im Schnitt alle zwei Tage ein beschädigtes Einsatzfahrzeug auf (Knoop & Röthmeier, 2012). Mitte letzten Jahres erschien eine unvollständige Aufzählung von Pressemeldungen, die berichtete Unfälle zwischen Januar 2016 und Anfang Juni 2017 auflistet (Herrnkind, 2017). Dort werden insgesamt 115 Meldungen (67 in 2016 und 48 im ersten Halbjahr 2017) genannt. In den Meldungen gibt es insgesamt 5 Unfälle mit Todesfolge (mit 7 getöteten Personen) und in 84 der Unfälle mit Verletzten wurden 198 Personen verletzt, mindestens 33 von ihnen schwer. Laut einer Schätzung soll etwa alle 272.000 Einsatzfahrten ein Mensch zu Tode kommen (Bockting, 2007). Bei einem geschätzten Einsatzaufkommen von 8.4 Millionen Einsatzfahrten im Jahr (Schmiedel, 2015), wären das 31 getötete Personen pro Jahr. In der Auflistung der Pressemeldungen wurden in 18 Monaten „nur“ sieben getötete Personen erwähnt. Somit scheinen weniger Todesfälle öffentlich zu werden oder tatsächlich seltener vorzukommen.

Aber jeder Unfall, der tatsächlich auf einer Einsatzfahrt mit Sondersignal passiert, ist in doppelter Hinsicht tragisch. Denn die Einsatzkräfte des Rettungsdienstes dürfen die Sondersignale nur nutzen, „wenn höchste Eile geboten ist, um Menschenleben zu retten oder schwere gesundheitliche Schäden abzuwenden“ (§35 StVO). Somit gibt es neben den nun zusätzlich verletzten Personen bereits eine Person, die auf schnelle Hilfe durch die Einsatzkräfte angewiesen ist und nun erst verzögert Hilfe bekommen kann. Zusätzliche indirekte Kosten ergeben sich daraus, dass evtl. ein anderes Einsatzfahrzeug zum eigentlichen Einsatzort

geschickt werden muss und für diesen Zeitraum nicht mehr verfügbar für andere Notfälle ist. Weiterhin muss das in den Unfall verwickelte Fahrzeug zunächst häufig aus dem Dienst genommen werden. Einen anschaulichen Überblick über die möglichen Nachwirkungen eines Verkehrsunfalls während einer Einsatzfahrt gibt Raheb (2011).

Mögliche Gründe für die erhöhten Unfallzahlen von Einsatzfahrern können die hohe Zahl an kritischen Fahrsituationen sein. Diese Situationen kommen während Einsatzfahrten etwa alle 19 Sekunden (Bockting, 2007; Pieper-Nagel & Wiegand, 2011; Ungerer, 2005; Unterkofler & Schmiedel, 1994) bis alle 3 Minuten vor (Deutscher Verkehrssicherheitsrat (DVR), 2008) und sind definiert als potenziell gefährliche Fahrsituationen, wie z.B. risikoreiche Fahrweise oder das Übergehen von Verkehrsgeboten (Unterkofler & Schmiedel, 1994). Zusätzlich sind Einsatzfahrer besonderen Anforderungen und Stressoren ausgesetzt (Deutscher Verkehrssicherheitsrat, 2008; Neukum, Lang & Krueger, 2003; Pieper-Nagel & Wiegand, 2011). Zu dem allgemeinen Verfolgen des Verkehrsgeschehens und besonders der notwendigen erhöhten Aufmerksamkeit aufgrund der eigenen Verkehrsüberschreitungen kommt der Einsatzstress hinzu. Das kann zum Beispiel ein Einsatzstichwort sein, das die Einsatzkraft an einen anderen Einsatz erinnert, oder die Kommunikation mit der Leitstelle oder dem Beifahrer. Unfälle mit Blaulichtbeteiligung<sup>1</sup> gehen zu fast zwei Drittel auf die Fahrer der Einsatzfahrzeuge zurück (Müller, 2010b) und diese sind es auch, die zentral bestimmen, wie sie unter Nutzung von Sondersignalen fahren. Die Einsatzfahrer als insgesamt überschaubarere Gruppe im Vergleich zu allen anderen Verkehrsteilnehmern sind daher eine Stellschraube, an der es sich lohnt anzusetzen, um die Verkehrssicherheit zu erhöhen.

Das Personal im Rettungsdienst kann aufgrund unterschiedlicher Ausbildungen als Fahrer von Einsatzfahrzeugen eingesetzt werden und die Voraussetzungen, um ein Rettungsfahrzeug (RTW) oder ein Notarzteinsatzfahrzeug (NEF) fahren zu dürfen, unterscheiden sich zwischen den Bundesländern. In Bayern beispielsweise muss ein Fahrer von Krankentransportfahrzeugen (KTW) bzw. RTW lediglich „geeignet“ sein, in Sachsen „fachlich geeignet“. Der Fahrer eines NEF muss in Bayern mindestens die Qualifikation zum Rettungssanitäter haben, jedoch darf der Notarzt ausnahmsweise auch selbständig fahren, wenn das Einsatzfahrzeug ansonsten nicht zum Einsatz kommen könnte (Art. 43, BayRDG). Typischerweise werden Rettungswagen in Deutschland von Rettungssanitätern gefahren, teilweise aber auch von medizinisch höher ausgebildeten Rettungsassistenten oder Notfallsanitätern.

Die unterschiedliche geforderte Qualifikation zeigt bereits, dass es nicht möglich ist, bei allen Fahrern von Einsatzfahrzeugen den gleichen Wissensstand vorauszusetzen oder festzustellen, wo angesetzt werden könnte, um alle Einsatzfahrer hinsichtlich der Verkehrssicherheit auf den gleichen Stand zu bringen. Obwohl die Ausbildung der Fahrer von Einsatzfahrzeugen eine der größten Präventionsmaßnahmen darstellt, um Unfälle zu vermeiden, gibt es weder deutschland- noch EU-weit einheitliche Vorgaben, wie diese Ausbildung auszusehen hat (Müller, 2010a; Albertsson & Sundström, 2011). Für Rettungssanitäter, die typischerweise als Einsatzfahrer eingesetzt werden, gibt es lediglich die Vorgabe, rechtliche Aspekte zu vermitteln, und auch im Praktikum, in dem sie typischerweise bei der Versorgung der Patienten eingesetzt werden, lernen die Rettungssanitäter nichts über das Führen der Einsatzfahrzeuge.

Im Januar 2013 wurde das Rettungsassistentengesetz durch das Notfallsanitätergesetz abgelöst. Für Rettungsassistenten und -assistentinnen (RA) in Deutschland galt: „Die Ausbildung soll [...] dazu befähigen [...] kranke, verletzte und sonstige hilfsbedürftige Personen [...] unter sachgerechter Betreuung zu befördern“ (§3 RettAssG). Die theoretische Ausbildung der Rettungsassistenten beinhaltete Themen wie Rettungsmittel/Rettungssysteme und Straßenverkehrsrecht, insbesondere Sonderrechte im Straßenverkehr, die durch den praktischen Einsatz vermittelt werden sollten (Anlage 1 A und §2, RettAssAPrV). Wie die Vermittlung des Wissens und besonders des Könnens auszusehen hat, wurde jedoch nicht festgelegt und konnte sich somit sehr stark zwischen den Ausbildungsstätten unterscheiden. Im

---

<sup>1</sup> Blaues Blinklicht laut StVO wird zur Vereinfachung im Folgenden als Blaulicht bezeichnet.

Ausbildungsziel der Notfallsanitäter kommt die Durchführung des Transportes gar nicht mehr vor, da sie für die Versorgung der Patienten zuständig sein sollen. Eine Ausbildung im Bereich Verkehrssicherheit muss daher für Notfallsanitäter nicht erfolgen, aber auch innerhalb der Ausbildung zum Rettungssanitäter gibt es kein einheitliches Vorgehen, wie das Thema Verkehrssicherheit vermittelt wird.

Auch aus der Perspektive der betrieblichen Arbeitssicherheit und des Arbeitsschutzes zeigt sich die Notwendigkeit das Thema Verkehrssicherheit im Rettungsdienst intensiver zu betrachten. Aus Daten der Unfallkasse des Bundes (Jost, 2012) geht hervor, dass 3% (2003) bzw. 5% (2008) der Arbeitsunfälle beim Deutschen Roten Kreuz (DRK) auf Verkehrsunfälle bei Einsatzfahrten zurückzuführen sind. Das ist eine verhältnismäßig kleine Zahl, z.B. im Vergleich zu Verletzungen durch Stolpern, Stürzen und Ausrutschen (2003: 38% bzw. 2008: 23%) oder Schnitte, Risse und Quetschungen (2003: 13% bzw. 2008: 16%). Die Arbeitsunfähigkeits-Tage (AU-Tage) durch diese Art von Unfällen ist jedoch deutlich höher: 2008 waren es 730 Tage durch Verkehrsunfälle bei Einsatzfahrten im Vergleich zu 83 Tagen durch Schnitte, Risse und Quetschungen und 495 Tage durch Stolpern, Stürzen und Ausrutschen. Auch die Zahlen von 2011 (Jost, 2012) zeigen die Tendenz, dass AU-Tage durch Verkehrsunfälle deutlich höher sind als beispielsweise durch Nadelstichverletzungen (Nadelstichverletzungen: 5% der Gesamtunfälle, 28 AU-Tage; Verkehrsunfälle bei Einsatzfahrten: 2.5% der Gesamtunfälle, 1412 AU-Tage).

In der Gesamtschau der Thematik ergibt sich damit ein dringender Bedarf, Fahrer von Einsatzfahrzeugen besser als bisher zu schulen und Rettungsdienstmitarbeiter auf diese Aufgabe vorzubereiten. Fahrsicherheitstrainings, die sich nur auf das Fahrzeughandling von Einsatzfahrzeugen beziehen, werden von unterschiedlichen Institutionen angeboten (z.B. „Sicherheitsprogramm Einsatzfahrzeuge“ des DVR) und im Internet finden sich Programme, die Szenarien für Übungsplätze vorschlagen (vgl. Unfallkasse Baden-Württemberg, 2010). Die Reaktionen anderer Verkehrsteilnehmer, die sehr unterschiedlich ausfallen und somit zu kritischen Situationen führen können, lassen sich auf einem Übungsplatz jedoch kaum trainieren. Im öffentlichen Straßenverkehr sind Übungen zudem kaum durchführbar, ohne dabei andere Verkehrsteilnehmer zu gefährden. Weiterhin sollte eine Inflation von Einsatzfahrten im öffentlichen Raum aufgrund von Übungsfahrten vermieden werden, da sie zu verringerter Rücksicht der anderen Verkehrsteilnehmer führen könnte. Abhilfe kann in diesem Fall ein simulatorgestütztes Training bieten, wie es vom DVR und von der Deutschen Gesetzlichen Unfallversicherung (DGUV) entwickelt wurde und das in dem vorliegenden Projekt evaluiert wird.

Hintergrund des Trainings sind Erkenntnisse zu Unfallursachen. Die allgemeine Statistik (Statistisches Bundesamt, 2017) zeigt, dass personenbezogene Fahrfehler im Vergleich zu technischen Mängeln und allgemeinen Ursachen (z.B. Straßen- oder Witterungsverhältnisse) der häufigste Grund für Verkehrsunfälle sind, insgesamt 88.1%. So gingen lediglich 7.7% der Unfälle mit Personenschaden auf allgemeine Ursachen zurück und bei 0.9% der beteiligten Fahrzeuge konnten technische Mängel als Ursache festgestellt werden, wobei diese etwa zur Hälfte auf Zweiräder entfallen (Statistisches Bundesamt, 2017). Jedoch ist auch bei diesen allgemeinen und technischen Gründen der Faktor Mensch zu berücksichtigen, so sind etwa technische Gründe oft auf mangelhafte Wartung eines Fahrzeugs zurückzuführen. Weiterhin wurde oben bereits beschrieben, dass bei Unfällen während Einsatzfahrten häufig die Fahrer der Einsatzfahrzeuge als Hauptverursacher ermittelt werden (Müller, 2010a, 2010b, Pieper-Nagel & Wiegand, 2011). In der Studie der BASt (Unterkofler & Schmiedel, 1994) wurden Unfallursachen bei Einsatzfahrten weiterhin hinsichtlich statistischer Kenngrößen untersucht. Dort wurde festgestellt, dass Unfälle häufig bei kleineren Geschwindigkeiten besonders im Kreuzungsbereich geschehen. Eine höhere Motorleistung gehe allerdings mit höheren Kollisionsgeschwindigkeiten einher. Weiterhin ist besonders die erste Einsatzfahrt der Schicht gefährdet und Anfahrten zur Einsatzstelle im Vergleich zu Krankenhaus oder Rückfahrten. Das Fahren nur mit Blaulicht jedoch ohne Einsatzhorn führt nachts ebenso wie bei Fahrern ohne Fortbildungsmaßnahmen über das Fahren von Einsatzfahrzeugen häufiger zu Unfällen.

Fahrsicherheitstrainings werden seit einigen Jahren in unterschiedlicher Form angeboten. In Deutschland gibt es beispielweise spezielle Zentren, in denen neben PKW- und Motorrad-Fahrern auch Berufskraftfahrer in Fahrsituationen wie Aquaplaning, Schleudern des Fahrzeugs oder Ausweichmanövern geschult werden können. Seit Einführung solcher Trainingsangebote beschäftigt sich auch die Forschung mit den Reaktionen von Teilnehmern und Auswirkungen der Trainings.

Ein umfangreiches und gut geeignetes Modell für vollständige Evaluationen von Trainings liefert Kirkpatrick (1998): Sein 4-Ebenen-Modell betont die Notwendigkeit Trainings auf vier Ebenen (Reaktion, Lernen, Verhalten und Resultate) zu evaluieren, da nur so wichtige und aussagekräftige Daten zur Wirksamkeit eines Trainings erfasst werden können (Kirkpatrick, 1996). Auf der ersten Ebene soll die „Reaktion“ der Teilnehmer auf das Training erfasst werden. Dabei werden Reaktion wie die Zufriedenheit mit dem Training und die Wahrnehmung, ob das Training als hilfreich oder sinnvoll erachtet wird, betrachtet. Diese über Befragung erhobenen Daten sollen Aufschluss darüber geben, wie hoch die Motivation und das Interesse der Teilnehmer für Lernen und eine Umsetzung des Gelernten sind. Die zweite Ebene betrifft das „Lernen“, d.h. eine veränderte Einstellung der Teilnehmer sowie den Wissenszuwachs zum Thema des Trainings. Um die Möglichkeit des Transfers in den Alltag respektive den Arbeitsplatz zu ermöglichen, sind Lerngewinn sowie die Einstellungsänderung notwendige Voraussetzungen. Die dritte Ebene, „Verhalten“, soll feststellen, inwieweit die Inhalte des Trainings einen Einfluss auf das reale (Arbeits-)Handeln haben. Es sollen also Verhaltensparameter untersucht werden, die auf unterschiedliche Art und Weise erfasst werden können. Neben Befragungen der Teilnehmer direkt oder deren Kollegen, können Verhaltensbeobachtungen oder -messungen stattfinden. Die vierte und letzte Evaluationsebene betrifft die Messung der Ergebnisse des Trainings und wird mit „Resultaten“ bezeichnet. Die Ergebnisse können dabei sehr unterschiedlich definiert werden, was die letzte Ebene schwieriger zu erfassen macht. Generell gilt, dass es sich bei Trainingsresultaten um die Ergebnisse des veränderten Erlebens oder Handelns der Trainingsteilnehmer handelt. Das New World Kirkpatrick Model (Kirkpatrick & Kirkpatrick, 2016) geht nicht von einem direkten Zusammenhang der Ebenen aus. Vielmehr können die Ebenen unterschiedliche Informationen über den Erfolg des Trainings bieten. Zeitlich gesehen kommen die Level 1 und 2 vor der 3 und 4, die sich auch gegenseitig bedingen können.



Abbildung 1: Neues Kirkpatrick Evaluationsmodell („New World Kirkpatrick Model“, Kirkpatrick & Kirkpatrick, 2016, übersetzt durch Autoren)

Bisher veröffentlichte Studien zur Wirkung von verschiedensten Fahrertrainings nutzen häufig Fragebögen oder Selbstevaluationen zu den Themen Geschwindigkeit, Abstand, Fahrtüchtigkeit, Risikowahrnehmung oder Fahrgewohnheiten (Molina, Sanmartin, Keskinen & Sanders, 2007; Perrissol, Smeding, Laumond & Le Floch, 2011; Rosenbloom, Shahar, Elharar & Danino, 2008; Tronsmoen, 2010). Molina et al (2007) betrachteten in ihren Fragebögen sehr junge Fahrer mit maximal drei Jahren Fahrerfahrung und fanden,

dass sich ein kombiniertes Training aus Simulatoren, Teststrecken und Gruppendiskussionen noch neun Monate später positiv auf die Selbstevaluation der Fähigkeiten für vorausschauendes („careful“) Fahren auswirkt. Tronsmoen (2010) befragte eine ähnliche Zielgruppe mit dem Ergebnis, dass Fahrer mit viel Training selbstkritischer waren und eine bessere Einstellung zur Verkehrssicherheit hatten. Perrissol et al. (2011) und Rosenbloom et al. (2008) befragten neben jungen auch ältere und erfahrenere Fahrer. Personen mit Verkehrsdelikten schätzten nach einem Training Situationen eher als potenziell gefährlich ein als vorher (Perrissol et al., 2011) und eine höhere Risikowahrnehmung nach einem Training mit theoretischen und praktischen Inhalten wurde bei allen Teilnehmern gefunden (Rosenbloom et al., 2008). Frauen und ältere Fahrer zeigten insgesamt eine höhere Risikowahrnehmung, bei jungen Fahrern zeigten sich entsprechend insgesamt höhere Trainingseffekte. Je nach Art der Fragestellungen werden anhand solcher Studien die ersten beiden Evaluationsebenen erfasst, die jedoch keine Aussage darüber erlauben, wie sich die Teilnehmer im Straßenverkehr tatsächlich verhalten. McKenna, Horswil und Alexander (2006) untersuchten zusätzlich zu Fragebögen anhand von Video-Sequenzen, wie sich die jungen Teilnehmer eines Gefahrenwahrnehmungs-Trainings mit maximal drei Jahren Fahrerfahrung in bestimmten Situationen verhalten würden. Vier verschiedene Arten von Videosequenzen und Fragestellungen wurden genutzt: 1) Aus Sicht des Fahrzeugführers sollten die Teilnehmer entscheiden, welche Geschwindigkeit sie in einer gezeigten Situation fahren würden. 2) Bei einem immer näher vorausfahrenden Fahrzeug sollte per Knopfdruck angegeben werden, ab wann die Teilnehmer sich unangenehm nah fühlten. 3) In einer Abbiegesituation sollte per Knopfdruck entschieden werden, bei welcher Lücke zwischen den Fahrzeugen sie zum Abbiegen ansetzen würden. 4) Bei verschiedenen Fahrszenen sollte während der Fahrt eine Bewertung abgegeben werden, ab wann diese eine potenziell kritische Situation darstellen. Es zeigte sich, dass die trainierte Gruppe weniger Risiko bei Geschwindigkeit und Abstand einging und schneller im Erkennen potenziell kritischer Situationen war. Auch diese Studie zielt dabei eher auf das Wissen als auf das tatsächliche Verhalten ab. Zu wissen, wie man sich in einer bestimmten Situation zu verhalten hat und das Risiko richtig einschätzen zu können, sind notwendige, jedoch keine hinreichenden, Voraussetzungen, sich angemessen zu verhalten. Zur Messung der dritten Ebene Verhalten bieten sich simulierte Fahrten an, besonders, wenn reale Fahrten aus den oben genannten Gründen nicht durchführbar sind. Solche Untersuchungen wurden beispielsweise von Damm, Nachtergaele, Meskali und Berthelon (2011) an jungen Männern sowie von Dorn und Barker (2005) an Polizisten im Vergleich zu einer Gelegenheitsstichprobe durchgeführt, um verschiedene Trainingsarten zu vergleichen. Damm et al. (2011) fanden dabei keinen Unterschied in Reaktionszeit und Anfangsgeschwindigkeit im Simulator, die trainierte Gruppe zeigte jedoch effizientere Ausweichmanöver. Bei den Polizisten zeigten sich weniger riskante Überholmanöver, eine verringerte Geschwindigkeit bei Fußgängern und besseres Fahren in der Mitte der Spur innerhalb der Stadt, als bei nicht polizeilich geschulten Teilnehmern (Dorn & Barker, 2005). Weiterhin untersuchten Isler, Starkey, Sheppard und Yu (2007) in einer kleinen Pilotstudie mit acht 16-Jährigen mit neuseeländischem Führerschein, wie gut die Messung realer Fahrten anhand von GPS-Daten erfolgen kann. Die Forscher nehmen später an, dass sich durch diese Messungen sichere von unsicheren Fahrern unterscheiden lassen (Isler, Starkey & Sheppard, 2011). Das Training von Isler et al. (2011) mit Inhalten zu Situationsbewusstsein, Gefahrenwahrnehmung und Risikomanagement führte bei den 15- bis 18-jährigen Teilnehmern zu einer sichereren Einstellung gegenüber dichtem Auffahren und Überholen und zu Verbesserungen bei Gefahrenwahrnehmung und Blickverhalten. Die vierte Ebene Resultate lässt sich im Straßenverkehr vor allem an der Anzahl und Schwere der Unfälle festmachen. Dies ist jedoch auch durch viele andere Faktoren beeinflusst (technische Entwicklungen, Assistenzsysteme, verbesserte Verkehrsführungen usw.). Auch andere denkbare Resultate-Parameter können aufgrund von Konfundierungen immer nur eine Annäherung geben, so dass eine langfristige Messung der trainierten Teilnehmer im Vergleich zu untrainierten Personen im Rahmen der Resultate-Messung sinnvoll wäre. Die bisherigen Studien zum Thema Fahrsicherheits-training zeigen, dass weder die Messung von Verhaltensänderungen im realen Straßenverkehr noch die Ebene der Trainings-Resultate hinreichend untersucht wurden.

Auch hinsichtlich des in dieser Studie zu evaluierenden Trainings „Verkehrssicherheit bei Einsatzfahrten“ wurden bereits einige Evaluationsschritte durchgeführt. Das simulatorgestützte Training „Verkehrssicherheit bei Einsatzfahrten“ des DVR und der DGUV beinhaltet besonders die Vermittlung von Kompetenzen im Bereich des Entscheidungsverhaltens, der Gefahrenpotenziale sowie der Wahrnehmungs- und Informationsverarbeitungsprozesse (Wiegand & Pieper-Nagel, 2011). In den theoretischen Teilen werden selbst erlebte kritische Situationen der Teilnehmer gesammelt und Ursachen von Unfällen besprochen. Dabei wird besonders auf das Risiko von Einsatzfahrten im Vergleich zu Normalfahrten eingegangen sowie die Rolle der Einsatzfahrer näher beleuchtet. Ein weiterer Theorieblock beschäftigt sich mit den rechtlichen Grundlagen zu Sonder- und Wegerechten. Beim Thema Entscheidungsverhalten werden Variablen besprochen, die den Anreiz und die Risikoeinschätzung beeinflussen können. Bei Gefahrenpotenzialen geht es um fahrphysikalische Aspekte, wie Geschwindigkeitswahlen, Anhaltewege, Aufprallgeschwindigkeiten und was diese für den menschlichen Körper bedeuten. Im letzten Theorieblock werden die Anzahl der wahrnehmbaren Informationen unter Stress, scharfes Sehen und die geschwindigkeitsabhängige Blickverengung besprochen sowie Möglichkeiten erarbeitet, die Wahrnehmungsleistung besonders bei Einsatzfahrten aufrecht zu erhalten. Die Vermittlung einiger höherer kognitiver Fähigkeiten mittels der theoretischen Blöcke des Trainings bietet eine Grundlage, das Gelernte in realen Situationen anzuwenden (Goode, Salmon & Lenné, 2013). Dementsprechend durchfahren alle Teilnehmer des Trainings jeweils zwei Einsatzfahrten am Simulator, um vermitteltes Wissen anzuwenden und für Stresssituationen während der Fahrt sensibilisiert zu werden. Die weiteren Teilnehmer können die Fahrten der anderen beobachten und werden angehalten auf bestimmte Aspekte zu achten, wie die Nutzung der Sondersignale, die Geschwindigkeit oder kritische Situationen während der Fahrt. So können im direkten Anschluss an die Fahrt durch Feedback zwischen den Teilnehmern Entscheidungsalternativen erarbeitet werden und gegenseitiges Lernen kann erfolgen. Die erste Simulatorfahrt ist eine Eingewöhnungsfahrt, die zweite Fahrt wird durch Zusatzaufgaben erschwert, die Stresssituationen, Ablenkungen oder Kommunikation mit der Leitstelle bzw. dem Beifahrer simulieren. Die Simulatorfahrt endet jeweils an vorgegebenen Orten oder bei Eintreten eines Unfalls, um diesen zu besprechen. Diese Art des Trainings bietet eine gute Möglichkeit des Transfers in ähnliche oder neue Situationen (Ivancic & Hesketh, 2000). Das Training wurde bereits im Rahmen seiner Entwicklung evaluiert. Die durchgeführten Befragungen auf der ersten Evaluationsebene zeigten bei den Teilnehmern eine sehr hohe Akzeptanz sowie eine durchschnittliche Bewertung des Trainings mit „sehr gut“ (Wiegand, 2011). Die Ergebnisse der Erfassung einiger Merkmale der zweiten Evaluationsebene stellte Varnskühler (2011) in seinem Vortrag auf der A+A 2011 vor. Sechs Monate nach der Durchführung gaben 29 (43%) Teilnehmer dem Training gute Noten in der Relevanz für die Arbeit. Zudem berichteten sie von einer verbesserten Einstellung zur Verkehrssicherheit und gaben an, jetzt mit gutem Beispiel voran zu gehen. In einem Vortrag auf der TTD Konferenz wurden weitere Evaluationsergebnisse zum Wissenszuwachs von 76 befragten Teilnehmern vorgestellt (Pieper-Nagel, 2011). Das Wissen um das Entstehen kritischer Situationen im Einsatz, der Risikoeinschätzung sowie der rechtlichen Grundlagen konnte bei dem zum großen Teil ehrenamtlichen Teilnehmern mit wenigen Einsätzen im Monat von einem mittleren Niveau ausgehend angehoben werden.

Ein weiterer Aspekt, der im Rahmen von Verkehrssicherheitstrainings zu beachten ist, ist die Risikohomöostase. Wilde (1998) beschreibt in seinem Modell eine persönliche Risikoschwelle zur Maximierung des Nutzens bei wahrgenommenen Kosten. Im Straßenverkehr steht diese Schwelle dem wahrgenommenen, realen Risiko gegenüber, und es wird versucht, Abweichungen zu minimieren. Sicherheitssysteme erhöhen das subjektive Sicherheitsgefühl und verringern somit das wahrgenommene Risiko. Die so entstehende Abweichung von der persönlichen Risikoschwelle wird durch schnelleres Fahren (und damit dem Nutzen des schnelleren Ankommens) ausgeglichen (Homöostase). In anderer Richtung können z.B. schlechte Sichtverhältnisse das subjektive Risiko erhöhen und somit zu vorsichtigerer Fahrweise führen. Fahrsicherheitstrainings, die lediglich auf Fahrzeughandling abzielen, erhöhen fahrerisches Können und häufig in noch größerem Maße das Selbstvertrauen, welches im Sinne der Homöostase die Risiko-

bereitschaft erhöht und damit in mehr Unfällen münden kann (Wilde, 1998). Daher sollte bei Trainings darauf abgezielt werden, die Risikobereitschaft zu verringern und die Einstellung zur Verkehrssicherheit zu verändern.

Zusammengefasst und in Übereinstimmung mit dem Konzept der Risikohomöostase bilanzieren Albertsson und Sundström (2011), dass sich Fahrsicherheitstrainings speziell für Fahrer von Rettungsfahrzeugen nicht allein auf das sichere Führen des Fahrzeuges richten, sondern auch Aspekte der Risikowahrnehmung, Unfallschwerpunkte und Selbsteinschätzung beinhalten sollten. Bisher wurden bei der Evaluation von Fahrsicherheitstrainings für verschiedene Zielgruppen keine Verhaltensveränderungen im realen Fahrgeschehen untersucht und auch Einstellungsänderungen wurden nur selten erfasst. Speziell für das Training „Verkehrssicherheit bei Einsatzfahrten“ des DVR und der DGUV wurden bisher Aspekte der ersten zwei Evaluationsebenen nach Kirkpatrick (1998) erfasst, die dritte und vierte Ebene wurde bisher nicht betrachtet, so dass bislang noch keine Aussage zur Wirksamkeit des Trainings auf das primäre Ziel der verbesserten Verkehrssicherheit getroffen werden kann. Diese Lücke wurde in dem vorliegenden Projekt geschlossen.

## 2. Forschungszweck/-ziel

Das übergeordnete Forschungsziel der vorliegenden Untersuchung besteht darin, das Training „Verkehrssicherheit bei Einsatzfahrten“ ganzheitlich anhand des Modells von Kirkpatrick (1998) auf den vier Ebenen Reaktion, Lernen, Verhalten und Resultate zu evaluieren. Besonderes Augenmerk liegt dabei auf den bisher nicht evaluierten Ebenen Verhalten und Resultate. Es soll die Einstellungs- und Verhaltensänderung der Einsatzfahrer des Rettungsdienstes sowie deren Beanspruchung durch Einsatzfahrten untersucht werden. Ausgehend von der Grundannahme, dass das Training einen positiven Effekt auf allen vier Evaluationsebenen hat, werden im Folgenden die einzelnen Annahmen für die Ebenen näher spezifiziert.

Hypothese 1 zur Reaktionsebene:

Die Teilnehmer nehmen das Training als sinnvoll und hilfreich wahr.

Wie bereits im 4-Ebenen-Modell beschrieben, ist die Reaktion der Teilnehmer auf das Training wichtig für die weiteren Veränderungen, die sich ergeben sollen. Wenn ein Training als sinnvoll erachtet wird und zusätzlich hilfreiche Handlungsstrategien vermittelt, besteht die Chance, dass die Teilnehmer ihre Einstellung hin zu sicherem Fahren ändern. Bisherige Evaluationsergebnisse des Trainings „Verkehrssicherheit bei Einsatzfahrten“ zu dieser Ebene waren vielversprechend, werden aber im Sinne der Generalisierbarkeit und Replizierbarkeit sowie zur Vollständigkeit der Wirksamkeitsdarstellung auch in der vorliegenden Studie durchgeführt.

Hypothese 2a zur Lernebene:

Die Teilnehmer lernen durch das Training relevantes Wissen zur Verkehrssicherheit.

Hypothese 2b zur Lernebene:

Die Teilnehmer verändern durch das Training ihre Einstellung hin zu einer höheren Bedeutung von Verkehrssicherheit, höherem Verantwortungsbewusstsein und geringerer Risikobereitschaft.

Eine positive Reaktion auf das Training, wie sie in Hypothese 1 postuliert wird, reicht allein nicht aus, um Verhaltensänderungen im Arbeitsalltag zu erreichen. Das auf der zweiten Evaluationsebene gemessene

Lernen ist eine notwendige Voraussetzung, um auch im Verhalten echte Änderungen hervorzurufen. Es werden zwei zentrale Lernbereiche in den Hypothesen angenommen: Wissen und Einstellungen. Wissen kann losgelöst und neben realem Handeln und Situationen ent- und bestehen, während sich Einstellungen durch Erfahrungen entwickeln und die Bereitschaft einer Person darstellen, auf einen Gegenstand, Person oder Situation in einer bestimmten Art zu kognitiv, affektiv und behavioral zu reagieren. Da das Training sowohl Wissen vermittelt als auch reale Erfahrung kritischer Verkehrssituationen im Simulator bietet, werden beide Ebenen evaluiert: Hypothese 2a betrachtet den Wissenszuwachs hinsichtlich Unfallentstehung, Rolle der Einsatzfahrer, rechtliche Grundlagen, Fahrphysik, Wahrnehmung und Risikoeinschätzung. Die Einstellungsänderung, die in Hypothese 2b betrachtet wird, beinhaltet die Erfassung von Einstellungen zur Verkehrssicherheit, verkehrsbezogenen persönlichen Merkmalen und Erhebung der Risikobereitschaft. Da Wissen Vergessenseffekten unterliegt und Einstellungen sich durch Erfahrungen ändern können, wird auch die Nachhaltigkeit der Änderungen durch eine spätere wiederholte Messung dieser Faktoren evaluiert.

Hypothese 3 zur Verhaltensebene:

Das Training hat einen positiven Effekt auf das reale Fahrverhalten der Teilnehmer.

Die Verhaltensebene betrachtet den zentralsten Evaluationsparameter, da das reale Fahrverhalten der Trainingsteilnehmer den unmittelbarsten Einfluss auf die Verkehrssicherheit bei Fahrten mit Sonder- und Wegerechten haben sollte. Ein Blick auf die Verkehrsunfallstatistik von 2016 zeigt, dass die vier häufigsten Unfallursachen im Fehlverhalten der Fahrzeugführer in Fehlern beim „Abbiegen, Wenden, Rückwärtsfahren, Ein- und Anfahren“, der Missachtung der Vorfahrt, Abstandsfehler sowie nicht angepasster Geschwindigkeit liegen (Statistisches Bundesamt, 2017). Dabei fallen nicht angepasste Geschwindigkeit sowie Abstandsfehler besonders auf jüngere Fahrer zurück, Abbiegefehler und Vorfahrtsmissachtungen werden häufiger durch ältere Fahrer verursacht (Statistisches Bundesamt, 2017). Besonders hohe Geschwindigkeiten stellen ein Risiko im Straßenverkehr dar, da sie mit höheren Aufprallgeschwindigkeiten und somit höherem Verletzungsrisiko einhergehen. Hinsichtlich der zu untersuchenden Fahrweise spielt auch das vorausschauende Fahren eine Rolle. Einige Forscher definieren die vorausschauende Fahrweise so, dass die Absicht anderer Fahrer vorhergesehen wird und so frühzeitig darauf reagiert werden kann, dabei ist der Abstand zu den vorausfahrenden Fahrzeugen ein wichtiger Sicherheitsfaktor, um Kollisionen zu vermeiden (Eissfeldt & Wagner, 2003; Van der Hulst, Meijman & Rothengatter, 1999). Eine genaue Darstellung der erfassten Fahrverhaltensvariablen findet sich in Abschnitt 0.

Hypothese 4a zur Resultateebene:

Das Training vermindert die erlebte Beanspruchung der Teilnehmer während der Einsatzfahrten.

Hypothese 4b zur Resultateebene:

Das Training hat einen positiven Effekt auf typische Kennwerte einer Einsatzfahrt.

Ähnlich wie bei der Lernebene können im Bereich der Resultate verschiedene Aspekte unterschieden werden. Die vierte Hypothese ist daher in zwei unabhängige Hypothesen unterteilt. Hypothese 4a betrachtet die erlebte Beanspruchung während einer Einsatzfahrt als mögliches Resultat, vor dem Hintergrund, dass besseres Wissen, veränderte Einstellungen und vorausschauenderes Fahrverhalten helfen, kritische Situationen zu vermeiden und damit den allgemeinen Stresspegel bei der psychischen Belastung durch Einsatzfahrten zu senken. Je nach individuellen Voraussetzungen kann psychische Belastung zu unterschiedlichen kurz- oder langfristigen psychischen Beanspruchungen führen. Dabei wird psychische Belastung als „Gesamtheit aller erfassbaren Einflüsse, die von außen auf den Menschen zukommen und



psychisch auf ihn einwirken“ und psychische Beanspruchung als „unmittelbare (nicht langfristige) Auswirkung der psychischen Belastung im Individuum in Abhängigkeit von seinen jeweiligen überdauernden und augenblicklichen Voraussetzungen, einschließlich der individuellen Bewältigungsstrategien“ definiert (DIN EN ISO 10075-1, 2000, S.3). Da die psychische Beanspruchung mit individuellen Voraussetzungen und Bewertungsprozessen zusammenhängt, kann diese mithilfe des Empfindens der Beteiligten sowie körperlichen Parametern erfasst werden. Kurzfristige negative Beanspruchungsfolgen wie Ermüdung, Stress, Monotonie- oder Sättigungsempfinden können bei häufigem Auftreten längerfristig in psychische und/oder somatische Gesundheitsprobleme übergehen (vgl. Glaser & Herbig, 2012; Herbig & Glaser, 2013), so dass eine Minderung der Beanspruchung durch das Training auch positive Resultate im Sinne des Arbeits- und Gesundheitsschutzes zeigen könnte. Hypothese 4b beinhaltet die Erfassung von Kenngrößen, die typischerweise mit Einsatzfahrten einhergehen. Eine wirtschaftliche Größe ist beispielsweise die Hilfsfrist, die in jedem Bundesland gesondert geregelt ist und aufgrund derer die ökonomische Verteilung von Rettungswachen erfolgt. Die Hilfsfrist ist die Zeit, die vom Eingang der Meldung bzw. von Fahrtbeginn bis zum Eintreffen am Unfallort verstreicht. Ein Unfall während einer Einsatzfahrt erhöht diese Zeit und stellt somit eine weitere mögliche wirtschaftliche Kenngröße dar, die anhand der Schwere und der Anzahl der Schadensfälle erfasst werden kann. Bei der Erfassung von Unfällen gilt es dennoch zu beachten, dass diese multikausal verursacht werden. So kann sich eine vorausschauende Fahrweise, z.B. aufgrund erhöhter Aufmerksamkeit oder geringerer Geschwindigkeit und somit mehr Zeit für Entscheidungen, positiv auf das Unfallgeschehen auswirken. Viele weitere Variablen, wie Wetter, Verkehrsdichte, Müdigkeit oder die Reaktion der anderen Verkehrsteilnehmer sind für das Unfallgeschehen relevant, aber nicht durch ein Training beeinflussbar. Im Rahmen der Studie müssen sie jedoch betrachtet werden. Sollte das Training zu einer vorausschauenderen und sichereren Fahrweise führen, könnte sich dies (mit der gebotenen Vorsicht im Hinblick auf andere Einflüsse auf diese Parameter) positiv auf diese Kennwerte auswirken.

#### Hypothese 5:

Erfahrung hat einen Einfluss auf den Trainingseffekt.

Zusätzlich zu den vier Evaluationsebenen werden Variablen erfasst, die auf mehreren Ebenen einen Einfluss haben können. Die Erfahrung der Einsatzfahrer stellt dabei ein wichtiges Maß dar, das auf die in den Hypothesen 1 bis 4b postulierten Trainingseffekte differenziell wirken kann. Erfahrung wird oft im Sinne der Zeit seit Beginn der Kompetenzentwicklung in einer Domäne angegeben, in Fall des Autofahrens häufig anhand der Jahre seit dem Besitz des Führerscheins. Abzutrennen ist hiervon der Begriff der Häufigkeit oder Intensität, die allgemein dem Erhaltens- oder Hinzulernen dient (vgl. Hacker & Skell, 1993; Hacker, 2005) und hier anhand der jährlichen Fahrleistung erfasst werden kann. Wie diese Punkte schon zeigen, ist Erfahrung ein außerordentlich komplexes Konstrukt (vgl. Ericsson et al., 2006) und kann im Kontext von Fahrfähigkeiten auch den Besitz des Führerscheins für verschiedene Fahrzeugklassen, Wissenszuwachs durch Trainings oder Erfahrung mit spezifischen Situationsklassen (z.B. Fahren bei extrem schlechten Witterungsbedingungen) enthalten. Vor diesem Hintergrund ist denkbar, dass auf der Ebene des Lernens für erfahrene Fahrer geringere Zuwächse zu beobachten sind als bei unerfahrenen, etwa weil bereits ein höheres Ausgangsniveau beobachtet werden kann, während gleichzeitig neues Fahrverhalten vielleicht schneller durch erfahrene Fahrer umgesetzt werden kann. Auch wenn eine gerichtete Hypothese in diesem Bereich kaum möglich ist, gibt es eine Reihe von Befunden, die die Bedeutung von Erfahrung belegen und ihre mögliche Bedeutung für Zielgruppe und etwaige Anpassung des Trainings verdeutlichen: Obwohl eine Einsatzfahrt nie wirklich zur Routine wird bzw. werden darf, kann es dennoch sein, dass Einsatzfahrer mit höherer Erfahrung gelassener mit verschiedenen Situationen, auch im Straßenverkehr, umgehen (Neukum et al., 2003). Müller (2003, 2011) beschreibt mangelnde Erfahrung als Risikofaktor für Unfälle bei Einsatzfahrten, die aufgrund fehlender Ausbildung zustande kommt. Laut Müller (2003) müssen gerade Fahrer von Einsatzfahrzeugen ein hohes Maß an fahrerischem Können aufweisen, um die mangelnde

Erfahrung der anderen Verkehrsteilnehmer zu kompensieren. Weiterhin geht Müller (2003) darauf ein, dass besonders unter Zeitdruck und Stress panische Fahrreaktionen auftreten können, da weniger Kapazitäten zur Informationsverarbeitung zur Verfügung stehen. Geringe Fahrerfahrung kann den Fahrstress noch zusätzlich erhöhen (Müller, 2003). Auch die BAST geht in ihrer Analyse auf die Erfahrung der Einsatzfahrer ein (Unterkofler & Schmiedel, 1994). Als Ursache für Unfälle besonders bei jüngeren Einsatzfahrern wurde eine hohe Risikofreude verknüpft mit mangelnder Fahrpraxis auf Einsatzfahrzeugen ermittelt. Dahingegen sind auch negative Effekte von Erfahrung denkbar – so könnte höhere Erfahrung dazu führen, dass z.B. aus Rücksicht häufiger nur mit Blaulicht und ohne Einsatzhorn gefahren wird (Unterkofler & Schmiedel, 1994). Krüger (2010) stellte in einem Vortrag die hohe physiologische Beanspruchung während Fahrten unter Nutzung von Sonder- und Wegerechten vor und zeigte, dass bei Novizen im Vergleich zu Experten deutlich höhere emotionale Beanspruchungen zu finden waren.

Neben den in Hypothesen aufgeführten Variablen ist anzunehmen, dass weitere Variablen den Erfolg des Trainings beeinflussen können. Daher werden auch eine Reihe von Kontroll- oder potenziell moderierenden Variablen erfasst. Das beschriebene Evaluationsvorgehen misst Teile eines sehr komplexen Systems, das durch viele Aspekte beeinflusst sein kann. Ein zu erfassender Faktor ist die Arbeitsorganisation bzw. die Arbeitsbedingungen. Nur wenn die Arbeitsbedingungen es ermöglichen, neu erworbenes Wissen auch umsetzen zu können, ist eine Verhaltensänderung möglich. Der Wissenserwerb durch ein Training hängt zudem maßgeblich vom bisherigen Wissensstand ab, der wiederum durch arbeitsorganisatorische Maßnahmen, wie Personalauswahl und -entwicklung, beeinflusst ist. Neben diesen allgemeinen arbeitsbezogenen Bedingungen sind die spezifischen Vorgaben zur Einsatzfahrt und alle den eigentlichen Einsatz betreffenden Informationen und Bedingungen zu erfassen.

Alle einzelnen Maße und Parameter der verschiedenen Evaluationsebenen werden im Abschnitt 3.2 vorgestellt und inklusive ihrer psychometrischen Eigenschaften beschrieben.

### 3. Methodik

#### 3.1 Studiendesign

Um die für die Evaluierung notwendige eindeutige Rückführung der Effekte auf das Training zu ermöglichen, wurde als erweitertes Interventions-/Kontrollgruppendesign ein Solomon-Vier-Gruppen-Versuchsplan gewählt, bei dem bei zwei Gruppen eine Vorhermessung und bei zwei Gruppen das Training durchgeführt wird. Die Versuchspersonen sollen in Abhängigkeit der Erfahrung und des Ortes der Rettungswache möglichst gleichmäßig auf die vier Gruppen verteilt werden. Damit handelt es sich bei der Erhebung um ein quasiexperimentelles Untersuchungsdesign. Die Gruppen 1 und 3 stellen dabei die Experimentalgruppen dar, bei den Gruppen 2 und 4 handelt es sich um die Kontrollgruppen (vgl. Tabelle 1 zur Definition der einzelnen Gruppen). Zur Erhöhung der Teilnahmemotivation und für eine erste Erfassung der Replizierbarkeit der Ergebnisse sind die Kontrollgruppen als Wartegruppen ausgelegt. Durch den Solomon-Versuchsplan sollten mögliche konfundierende oder Störvariablen zusätzlich zu Zeit- und Testeffekten (etwa eine sozial erwünschte Antwortstrategie bei Nachhermessung nur dadurch, dass die Themen bei der Vorhermessung bekannt wurden) kontrolliert werden. Mit der Auslegung der Kontrollgruppen als Wartegruppen und einer „Nachhaltigkeitserfassung“ zur Erhebung längerfristiger Trainingseffekte ergeben sich bis zu vier Messzeitpunkte pro Person, die wiederum mehrere Schichtmessungen umfassen (vgl. Abbildung 2). Dieses Design und das gesamte Vorgehen wurden der Ethikkommission bei der Medizinischen Fakultät der LMU München unter dem Kennzeichen 206-14 vorgelegt, es wurden keine Einwände erhoben.

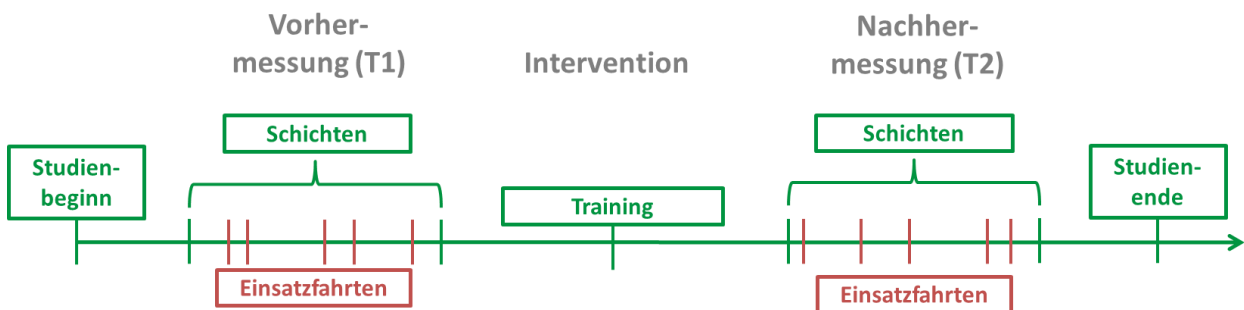


Abbildung 2: Datenerhebung – Ablauf für Interventionsgruppen mit Vorhermessung (nur T1/T2-Messung)

Die geplante Stichprobengröße lag bei insgesamt 200 Personen, um alle vier Versuchsgruppen mit einer angestrebten Power von .80 vergleichen zu können. Schon recht frühzeitig im Projekt wurde jedoch absehbar, dass die Probandenakquise deutlich schwieriger als erwartet und eine entsprechende Gruppengröße im geplanten Untersuchungszeitraum nicht zu erreichen war, so dass die Gruppe 4 des Designs fallengelassen wurde und damit insgesamt nur drei Versuchsgruppen in die Studie eingehen (vgl. Tabelle 1).

Darüber hinaus wird darauf verzichtet, alle potenziellen Kontrollvariablen mit Relevanz für die Verhaltens- und Resultateebene in den Analysen mitzuführen, da die Ergebnisse aufgrund der hohen Variabilität der erfassten Schichten dann kaum noch interpretierbar wären (vgl. auch Abschnitt 0). Stattdessen werden Einflussfaktoren auf die Beanspruchung von Mitarbeitern im Rettungsdienst (vgl. Abschnitt 7.4) und Einflussfaktoren auf Sondersignalfahrten (vgl. Abschnitt 7.1 und 7.2) als generelle Fragen unabhängig vom Evaluationsdesign untersucht. Bei letzteren Auswertungen werden zudem vertiefte Analysen über (potenziell) kritische Situationen bei Einsatzfahrten vorgestellt (vgl. Abschnitt 7.3).

Tabelle 1: Versuchsplan nach der Anpassung

	T1-(Prä)-Messung	Training	T2-(Post)-Messung	Training	T3-(Post)-Messung	T4-(Follow-up)-Messung
Gruppe 1	Y1	X1	Y2	/	/	Y4
Gruppe 2	Y1	/	Y2	X1	Y3	Y4
Gruppe 3	(Y1)	X1	Y2	/	/	Y4

Anmerkung: Y beinhaltet jeweils alle Messungen zu den jeweiligen Zeitpunkten T1-T4, in jedem ist ein T1-T4-Fragebogen, sowie Schichtmessungen (Ausnahme, wenn es in Klammern steht) enthalten. Der Wissenstest wird nur zu T1 und T2 durchgeführt, die direkte Trainingsevaluation jeweils nach dem Training X.

### 3.2 Datenerhebungsmethoden

Die Studie wurde multimethodal durchgeführt. Zur Übersicht zeigt Abbildung 3, wie in den folgenden Abschnitten die Darstellung der vielfältigen Methoden erfolgt. Alle Evaluationsebenen sowie weitere Einflussfaktoren wurden mindestens mittels Fragebogendaten erfasst und werden pro Ebene und inklusive Methodenentwicklung und Erhebungszeitpunkten berichtet. Für die Lernebene wurde zusätzlich ein Wissenstest entwickelt. Auf der Verhaltensebene wurden Fahrprofilaten, Videodaten und einige fragebogen-gestützte Schichtaspekte sowie Beifahrerbewertungen des Fahrstils erhoben. Hinsichtlich der Resultatebene wurden neben Fragebogendaten Schichtdaten erhoben sowie EKG-Messungen durchgeführt.

Ebene	Inhalt	Abschnitt Fragebogendaten	Sonstige Daten (Abschnitt)
Reaktion	Akzeptanz und Zufriedenheit mit einem Training	3.2.1.1	/
Lernen	Veränderte Einstellungen und Wissenszuwachs	3.2.1.2	Wissenstest (3.2.2)
Verhalten	Einfluss auf reales Handeln	3.2.1.3	Fahrprofilaten & Videodaten (3.2.3)
Resultate	Konsequenzen veränderten Verhaltens	3.2.1.4	Schichtdaten & EKGs (3.2.4)
Einflussfaktoren	Mögliche Kontrollvariablen	3.2.1.5	/

Abbildung 3: Überblick der erhobenen Daten mit Angabe der Abschnitte, in denen die Methoden beschrieben werden

#### 3.2.1 Fragebogendaten

Zu allen Messzeitpunkten wurde jeweils ein größerer Fragebogen durchgeführt. Der Eingangsfragebogen enthielt im Gegensatz zu den anderen Bögen mehr Skalen, da hier auch arbeitsbedingungsbezogene Variablen einmalig erhoben wurden. Weiterhin gab es eine Trainingsevaluation direkt nach dem Training sowie 1-2 Monate später und im Follow-up. Der Wissenstest wurde insgesamt zweimal durchgeführt, zu den Messzeitpunkten T1 und T2. Im Folgenden werden alle Fragebogendaten in der Reihenfolge der Evaluationsebene vorgestellt.

### 3.2.1.1 Reaktionsebene

Zur Erfassung der Reaktion auf das Training wurde direkt nach dem Training ein Fragebogen durch die Teilnehmer ausgefüllt. Dieser ist neu entwickelt, jedoch angelehnt an den nicht öffentlichen Evaluationsbogen des Workshops „Sichere Einsatzfahrt“ des UK Bundes Minas. Es wurden die vier Skalen Trainingsinhalte, Trainer, Rahmenbedingungen sowie Anwendbarkeit mittels 29 5-stufigen Items ermittelt (z.B. Ich bin motiviert, das Gelernte in realen Einsatzfahrten umzusetzen.). Weiterhin sollten die Teilnehmer angeben, ob sie das Training weiterempfehlen würden sowie dem Training eine Gesamtnote (6-stufig) geben und hatten Platz für freie Kommentare.

Tabelle 2: Interne Konsistenzen der Trainingsevaluationsbögen

Name der Skala	Beispielitem	Güte der Skala
Inhalte	Die Inhalte des Trainings haben meinen Erwartungen entsprochen	$\alpha=.90$
Trainer	Der Trainer hat die Trainingsinhalte verständlich vermittelt	$\alpha=.93$
Rahmenbedingungen	Der Austausch mit den Kollegen/Kolleginnen war hilfreich für mich	$\alpha=.67$
Anwendbarkeit	Ich finde das Training hilfreich für die Bewältigung von Einsatzfahrten	$\alpha_{\text{direkt}}=.93$ $\alpha_{\text{später}}=.93$ $\alpha_{\text{langfristig}}=.91$

Die später erfolgten Nachevaluationen 1-2 Monate später sowie zur Follow-up-Messung enthielt noch neun Items zur Erfassung der Anwendbarkeit (5 Items), erfolgte Weiterempfehlung (1 Item), Arbeitsrelevanz (1 Item) (jeweils 5-stufig) sowie erneut eine Schulnote und Platz für freie Kommentare. Alle Skalen weisen eine hohe Güte auf und konnten somit wie geplant genutzt werden.

### 3.2.1.2 Lernebene

Zur Erfassung der zweiten Evaluationsebene wurden sowohl ein selbstentworfener Wissenstest zur Messung des Wissenszuwachses verwendet als auch standardisierte Bögen, die verkehrsbezogene Persönlichkeitsmerkmale und die Einstellung zur Verkehrssicherheit erfassen. Der Wissenstest wird in Abschnitt 3.2.2 genauer beschrieben. Alle einstellungsbezogenen Variablen wurden zu allen vier Messzeitpunkten erfragt, um Änderungen über die Zeit überprüfen zu können.

Der Test zur Erfassung *verkehrsrelevanter Persönlichkeitsmerkmale*, kurz TVP (Spicher & Hänsgen, 2003), wurde auszugsweise genutzt. Es wurden insgesamt sechs Skalen (Extraversion (11 Items), Emotionale Labilität (4 Items), Gewissenhaftigkeit (3 Items), Verträglichkeit (5 Items), Bagatellisierung (5 Items) und Reaktanz (6 Items)) mit einem 4-stufigen Antwortformat („trifft nicht/nie zu“ bis „trifft genau/immer zu“) erfasst. Die Probanden sollten angeben, wie sehr Aussagen wie „Je schwieriger die Verkehrssituation ist, desto mehr macht mir das Fahren Spaß“ auf sie zutreffen. Da die Skalen nur auszugsweise genutzt wurden, wurden sie mittels interner Konsistenzen und Faktorenanalysen hinsichtlich Stabilität und Zuverlässigkeit überprüft. Da sich hierbei einige psychometrische Probleme zeigten, wurden teilweise neue Skalen erstellt und teilweise Items anderen Skalen zugeordnet. Für einige Items fand sich auch eine hohe Instabilität in der Bewertung und Skalenzuordnung und sie wurden nicht in die Analysen eingeschlossen. Dennoch konnte zum Teil nur eine mittelmäßige bis unzureichende Güte (z.B. interne Konsistenz zum dritten Messzeitpunkt bei Reaktanz) hergestellt werden. Tabelle 3 zeigt eine Auflistung der neu erstellten Skalen.

Tabelle 3: Neue Skalenstruktur der verkehrsbezogenen Persönlichkeitseigenschaften

Name der Skala	Beinhaltete Items	Güte der Skala
Extraversion	<p>Ich probiere gern mal aus, was in meinem Auto an Leistung drinsteckt.</p> <p>Wenn die Straße frei ist, drücke ich aufs Gas.</p> <p>Auf einer freien Strecke probiere ich gerne aus, wie gut ich Auto fahren kann.</p> <p>Es reizt mich, schneller als andere zu fahren.</p> <p>Ich nutze die Möglichkeiten meines Fahrzeuges voll aus.</p> <p>Wenn ich Auto fahre, möchte ich anders sein als die anderen, und nicht so ein gewöhnlicher Fahrer.</p> <p>Auf schwierigen Strecken, z.B. auf kurvigen Landstraßen, zeige ich anderen gern, wie man fährt.</p>	<p>Aufgeklärte Varianz: 46.1%</p> <p><math>\alpha_{T1}=.80</math>  <math>\alpha_{T2}=.82</math>  <math>\alpha_{T3}=.85</math>  <math>\alpha_{T4}=.87</math></p>
Sensation Seeking	<p>Je schwieriger die Verkehrssituation ist, desto mehr macht mir das Fahren Spaß.</p> <p>Ich liebe es, wenn der Adrenalinspiegel beim Autofahren steigt.</p> <p>Überraschungen im Straßenverkehr machen mir nichts aus.</p> <p>Ich fühle mich auch in schwierigen Verkehrssituationen sicher.</p>	<p>Aufgeklärte Varianz: 46.4%</p> <p><math>\alpha_{T1}=.60</math>  <math>\alpha_{T2}=.49</math>  <math>\alpha_{T3}=.55</math>  <math>\alpha_{T4}=.45</math></p>
Gewissenhaftigkeit	<p>Wenn ich keine Vorfahrt habe, schaue ich zweimal nach links und rechts, bevor ich losfahre.</p> <p>Bei einem Auto ist für mich in erster Linie die Sicherheit wichtig.</p> <p>Vor jeder Autofahrt kontrolliere ich die Bremsen.</p>	<p>Aufgeklärte Varianz: 54.5%</p> <p><math>\alpha_{T1}=.58</math>  <math>\alpha_{T2}=.48</math>  <math>\alpha_{T3}=.33</math>  <math>\alpha_{T4}=.61</math></p>
Unverträglichkeit	<p>Ich bin schon einmal Auto gefahren, obwohl ich mich eigentlich zu müde dazu fühlte.</p> <p>Auf der Autobahn bin ich schon einmal schneller als erlaubt gefahren.</p> <p>5-10 km/h über der Geschwindigkeitsbegrenzung zu fahren ist in Ordnung, weil dies jeder tut.</p> <p>Ich zeige anderen Autofahrern mit Handzeichen (z.B. Vogel zeigen), dass ich sie für unfähig halte.</p> <p>Andere Autofahrer fahren oft so miserabel, dass ich mich ärgere.</p>	<p>Aufgeklärte Varianz: 49.0%</p> <p><math>\alpha_{T1}=.74</math>  <math>\alpha_{T2}=.69</math>  <math>\alpha_{T3}=.70</math>  <math>\alpha_{T4}=.73</math></p>
Reaktanz	<p>Ich fahre Auto wie ich will, und niemand hat mir vorzuschreiben wie.</p> <p>Ich bin dagegen, dass der Staat neue Methoden als Sicherheitskontrollen für Fahrzeuge einführt.</p> <p>Wenn man vorsichtig fährt, ist der Sicherheitsgurt nicht nötig.</p> <p>Ich trage den Sicherheitsgurt nur, weil es gesetzlich vorgeschrieben ist.</p> <p>Ich bin der Meinung, dass Strafen für Verkehrsvergehen nicht härter werden sollten.</p>	<p>Aufgeklärte Varianz: 37.9%</p> <p><math>\alpha_{T1}=.53</math>  <math>\alpha_{T2}=.47</math>  <math>\alpha_{T3}=.15</math>  <math>\alpha_{T4}=.53</math></p>

Anmerkung: Güte enthält die aufgeklärte Varianz zum Zeitpunkt T1 berechnet mittels Faktorenanalyse, sowie die internen Konsistenzen (Cronbachs Alphas) zu allen vier Messzeitpunkten

Als weiteres Maß der Einstellungen wurde die *Risikowahrnehmung* im Straßenverkehr erfasst. Mittels dem Risk Perception Questionnaire (Rosenbloom et al., 2008) sowie Fragen zur Regelmissachtung (Delhomme, 1991) sollten die Probanden insgesamt 30 Fahrsituationen hinsichtlich ihres Risikos einschätzen. Dabei sollten sie auf einem 5-stufigem Antwortformat von überhaupt nicht riskant bis sehr riskant bewerten.

Beide Grundfragebögen gehen von Eindimensionalität aus, haben jedoch insgesamt 7 Items gemeinsam. Daher wurden auch diese Skalen hinsichtlich Stabilität und Zuverlässigkeit untersucht und neuen Skalen zugeordnet. Es entstanden insgesamt vier neue Skalen mit zufriedenstellenden psychometrischen Kennwerten: Risikowahrnehmung bei schweren, mittelschweren und leichten Regelverstößen sowie bei Normalfahrten. Tabelle 4 zeigt die neu entstandene Struktur mit den enthaltenen Items.

Tabelle 4: Neue Skalenstruktur der verkehrsbezogenen Risikowahrnehmung

Name der Skala	Beinhaltete Items	Güte der Skala
Risikowahrnehmung bei schweren Regelverstößen	Autofahren nach dem Konsum von drei Bieren a 0,5 l. Schneller als 130 km/h auf einer Bundesstraße fahren. 120 km/h auf einer Bundesstraße fahren. Überholen, wenn Sie durch einen LKW keine gute Sicht auf den Gegenverkehr haben. Überholen, obwohl man bereits denkt, dass es knapp wird. Eine rote Ampel überfahren. Auf offener Straße beschleunigen wenn man gerade überholt wird. Autofahren nach dem Konsum von zwei Bieren a 0,5 l. Einen anderen Fahrer beim Einscheren schneiden.	Aufgeklärte Varianz: 25.7%  $\alpha_{T1}=.62$ $\alpha_{T2}=.73$ $\alpha_{T3}=.60$ $\alpha_{T4}=.64$
Risikowahrnehmung bei mittelschweren Regelverstößen	Eine durchgezogene weiße Linie überfahren. Die Linie an einem Stoppschild überfahren, ohne komplett zum Stehen zu kommen. Häufiger Spurwechsel bzw. Slalom zwischen Fahrzeugen fahren, um schneller anzukommen. Beschleunigen, wenn die Ampel auf gelb schaltet. In der falschen Richtung in einer Einbahnstraße fahren (im Vorwärts- oder Rückwärtsgang). Fahren während eines wichtigen Telefongesprächs ohne Freisprechanlage.	Aufgeklärte Varianz: 42.8%  $\alpha_{T1}=.73$ $\alpha_{T2}=.74$ $\alpha_{T3}=.74$ $\alpha_{T4}=.66$
Risikowahrnehmung bei leichten Regelverstößen	110 km/h auf einer Bundesstraße fahren. Während des Fahrens Radio, Navigationsgerät oder Handy bedienen. Autofahren mit lauter und aufregender Musik im Hintergrund. Innerorts mit 60 km/h fahren. Ein langsames Auto rechts überholen, wenn es links nicht überholt werden kann. Nach einer schlaflosen Nacht fahren.	Aufgeklärte Varianz: 41.2%  $\alpha_{T1}=.71$ $\alpha_{T2}=.72$ $\alpha_{T3}=.58$ $\alpha_{T4}=.67$
Risikowahrnehmung bei Normalfahrten	Fahren während eines wichtigen Telefongesprächs mit Freisprechanlage. Während des Fahrens essen. Einen steilen Hang in einem hohen Gang hinunterfahren. Ein Fahrzeug mit zu niedrigem Reifendruck auf einer trockenen Fahrbahn fahren. Innerorts mit 50 km/h fahren. Fahren mit nur einer Hand am Lenkrad.	Aufgeklärte Varianz: 39.5%  $\alpha_{T1}=.69$ $\alpha_{T2}=.71$ $\alpha_{T3}=.62$ $\alpha_{T4}=.62$

Als weiterer Indikator für Einstellung wurde die *Self-Control-Skala* herangezogen (Grasmick et al., 1993, in deutscher Fassung von Seipel, 2012), die risikoreiches Verhalten unabhängig vom Straßenverkehr erfasst. Die Probanden sollten Aussagen wie „Hin und wieder setze ich mich kleinen Risiken aus, um mich auszu-

testen“ mittels eines 5-stufigen Antwortformates dahingehend bewerten, wie stark sie auf sich selbst zutreffen („trifft überhaupt nicht zu“ bis „trifft voll und ganz zu“). Die vier Items zeigten eine stabile und zuverlässige Datenstruktur (Cronbachs  $\alpha$  zwischen .83 und .89) und wurden daher originär beibehalten.

Als subjektives Maß für die *Einstellung zur eigenen Fahrfähigkeit* wurden die Probanden mittels drei Items des Driving Skill Questionnaire (Horswill, Waylen & Tofield, 2004) befragt. Zusätzlich wurden diese Items noch einmal angepasst auf den Rettungsdienst. So sollten die Probanden ihre Fahrfähigkeit nicht nur im Vergleich zu anderen Autofahrern, sondern auch zu anderen Rettungsdienstfahrern einschätzen. Es wurde eine 11-stufige Skala von „viel geringer/schlechter als der Durchschnitt“ bis „viel höher/besser als der Durchschnitt“ genutzt. Die Teilnehmer sollten in diesem Rahmen auch einschätzen, wie hoch sie die Wahrscheinlichkeit einschätzen, selbst in einen Unfall verwickelt zu werden und zwar jeweils im Vergleich zu anderen Autofahrern und zu anderen Rettungsdienstfahrern. Somit ergaben sich insgesamt vier Skalen, die für die Auswertungen genutzt werden können. Die selbsteingeschätzte Fahrkompetenz im Vergleich zur Allgemeinbevölkerung (Cronbachs  $\alpha$ =.85) bzw. zu anderen Rettungsdienstfahrern (Cronbachs  $\alpha$ =.89) sowie je als Einzelitems die wahrgenommene Unfallwahrscheinlichkeit im Vergleich zur Allgemeinbevölkerung und im Rettungsdienst.

Als letztes Maß für Einstellungen wurden die Probanden gefragt, wie gern sie selbst bzw. als Beifahrer unter Einsatz von Sonder- und Wegerechten fahren und wie sehr sie Sondersignalfahrten beanspruchen sowohl als Fahrer als auch als Beifahrer. Diese Fragen waren selbst entworfen und wurden mittels einer 5-stufigen Skala zu allen Messzeitpunkten erfasst.

### 3.2.1.3 Verhaltensebene

Die Verhaltensebene wurde größtenteils objektiv erfasst und wird genauer in Abschnitt 3.2.3 erläutert. Dennoch gab es zu allen gemessenen Schichten in allen Messblöcken einen Schichtfragebogen, der zusätzliche Maße für die Messung des Verhaltens erfasste. So wurden die Teilnehmer zum Ende der Schicht jeweils gebeten, auf einer 10-stufigen Skala anzugeben, wie sicher, vorausschauend, rasant bzw. hektisch sie über die Schicht hinweg gefahren sind. Zusätzlich bekamen die Beifahrer der Schicht dieselben vier Fragen und sollten den Fahrer bewerten. Da die Beifahrer versehentlich eine 11-stufige Skala erhielten, wurden gemäß dem von Cohen et al. (1999) empfohlenen Vorgehen beide Skalen auf die Prozentwerte maximal möglicher Punkte umgerechnet, um sie direkt vergleichen zu können. Somit kann das Verhalten sowohl subjektiv als auch aus der „Außenperspektive“ (d.h. etwas objektiver) erfasst werden. Weiterhin sollten die Einsatzfahrer angeben, ob sie während der Schicht ein kritisches Ereignis im Straßenverkehr erlebt haben und es ggf. kurz beschreiben.

### 3.2.1.4 Resultateebene

Für die Resultateebene wurden sowohl kurzfristige als auch längerfristige subjektive Maße erfasst. Für die Messung *längerfristiger Beanspruchungsfolgen* wurden zu allen vier Messzeitpunkten standardisierte, etablierte Skalen verwendet. Die Irritationsskala zur Erfassung arbeitsbezogener Beanspruchungsfolgen (Mohr, Rigotti & Müller, 2005) erfasst mittels acht Items kognitive (z.B. „Es fällt mir schwer, nach der Arbeit abzuschalten“) und emotionale Irritation (z.B. Ich reagiere gereizt, obwohl ich es gar nicht will.“). Auf einer 7-stufigen Skala von „trifft überhaupt nicht zu“ bis „trifft fast völlig zu“ sollten die Probanden bewerten, wie sich Belastungen in der Arbeit bei ihnen auswirken. Die beiden Skalen wie auch die Gesamtskala weisen sehr gute interne Konsistenzen auf (Cronbachs  $\alpha$  zwischen .82 und .94).

Weiterhin wurden mittels des Gießener Beschwerdebogen (Brähler, Hinz & Scheer, 2008) körperliche Beschwerden auf einer 5-stufigen Skala von keine bis starke Beschwerden erfasst. Die 24 Items können neben dem Gesamtbeschwerdedruck die vier Skalen Erschöpfungsneigung, Magenbeschwerden, Glieder-



schmerzen und Herzbeschwerden erfassen, zu denen es bevölkerungsrepräsentative Vergleichsstichproben gibt. Die Skalen weisen auch in unserer Stichprobe eine gute bis sehr gute Reliabilität auf (Cronbachs  $\alpha$  zwischen .83 und .96).

Zwei weitere langfristige Beanspruchungsmaße, die zu allen vier Messzeitpunkten erfasst wurden, sind der WHO-5 (Bech, 2004), der sowohl als Maß des Wohlbefindens in den letzten zwei Wochen als auch als Depressionsscreening (vgl. Henkel et al., 2003; Topp et al., 2015) eingesetzt werden kann, sowie das Maslach Burnout Inventory (MBI; dt. Büssing & Perrar, 1992) zur Messung von Burnout. Der WHO-5-Fragebogen erfasst positiv formuliert affektive und motivationale Hauptsymptome von Depressionen; positive Stimmung, Antrieb und Interesse und erlaubt mittels eines Cut-Off-Wertes von 13 ein Screening auf Depression, wobei ein positiver Prädiktionswert von etwa 34 darauf verweist, dass etwa ein Drittel aller Personen mit positivem Screening tatsächlich zum Zeitpunkt der Erhebung an einer Depression leiden (vgl. z.B. Henkel et al., 2003; Schneider et al., 2010). Er besteht aus fünf 6-stufigen Items (z.B. „In den letzten zwei Wochen war ich froh und guter Laune“) und zeigt sehr gute Reliabilität (Cronbachs  $\alpha$  = .86). Auch die beiden Burnout-Skalen emotionale Erschöpfung und Zynismus des MBI weisen sehr gute Reliabilitäten auf (Cronbachs  $\alpha$  zwischen .80 und .87).

Um *kurzfristige Beanspruchungsfolgen* erfassen zu können, wurde drei Skalen in alle Schichtmessungen zu allen Messzeitpunkten aufgenommen. Jeweils zu Schichtbeginn und -ende wurde die Basler Befindlichkeitskala (BSS, Hobi, 1985) mit ihren vier Subskalen Vitalität, Intrapsychischer Gleichgewichtszustand, Soziale Extravertiertheit und Vigilanz/kognitive Steuerung erfasst. Dieses Verfahren ist zur Veränderungsmessung geeignet und gut validiert. Anhand von bipolaren Adjektivlisten mit 16 Items werden die Bereiche Vitalität (z.B. kraftlos-tatkräftig), intrapsychischer Gleichgewichtszustand (z.B. ruhig-nervös), soziale Extravertiertheit (z.B. zurückgezogen-gesellig) und Vigilanz (z.B. wachsam-zerstreut) mit einer siebenstufigen Antwortskala erfasst. Im Hinblick auf kognitive Arbeitsanforderungen und Schichtarbeit sollten v.a. Aufmerksamkeit (Vigilanz) und Ermüdung (Vitalität) zur Erfassung kurzfristiger Beanspruchungsfolgen eine Rolle spielen. Veränderungen im intrapsychischen Gleichgewichtszustand und der sozialen Extravertiertheit sind vermutlich eher mit den emotionalen und sozialen Belastungen der Tätigkeit assoziiert. Zur Messung der Schläfrigkeit wurde zu Beginn und Schichtende sowie dreimal während der Schicht die Karolinska Sleepiness Scale (Åkerstedt & Gillberg, 1990) gemessen. Die Probanden sollen mithilfe einer verankerten 9-stufigen Skala von „sehr wach“ bis „sehr müde, große Probleme, wach zu bleiben, mit dem Schlaf kämpfend“ ihre momentane Müdigkeit einschätzen. Ebenfalls dreimal während der Schicht, als auch zum Schichtende sollten die Probanden zusätzlich den NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988) ausfüllen. Der NASA-TLX umfasst sechs bipolare kontinuierliche Dimensionen (drei aufgabenbezogene Skalen: Geistige Anforderungen, körperliche Anforderungen, zeitliche Anforderungen; eine verhaltensbezogene Skala: Leistung; zwei individuumsbezogene Skalen: Anstrengung, Frustration), die jeweils anhand einer zwanzigstufigen Skala (Wertebereich: 0-100) durch den Mitarbeiter eingeschätzt werden. Die Sensitivität und Kriteriums-Validität des Verfahrens ist gegeben: Die Ergebnisse korrelierten in Studien (z.B. Müller, Petru, Seitz, Englmann & Angerer, 2010; Müller, Petru & Angerer, 2011) erwartungsgemäß mit objektiven Kriterien, wie Aufgabenschwierigkeit, zeitlichen Anforderungen und erforderlichem Aktivitätslevel.

### 3.2.1.5 Kontrollvariablen

Da Arbeitsbedingungen Einfluss auf das Lernen und Verhalten von Mitarbeitern haben können, wurden diese zur Kontrolle sowie zu Ableitung potenzieller Handlungsempfehlungen für die Organisationen im Eingangsfragebogen mit etablierten Verfahren erhoben. Das Tätigkeits- und Arbeitsanalyseverfahren (TAA, Büssing & Glaser 2002) wurde in adaptierter Fassung für den Rettungsdienst genutzt, um arbeitsbezogene Anforderungen, Ressourcen und Stressoren in den Bereichen kognitive Anforderungen (4 Items), Lernerfordernisse (3 Items), Qualifizierungsmöglichkeiten (3 Items), zeitliche Transparenz/Vorhersagbarkeit (4

Items), Vorgesetztenfeedback (2 Items), Zeitautonomie (4 Items), Tätigkeitsspielraum (Handlungs- und Gestaltungsspielraum, 5 Items), Partizipationschancen (4 Items), räumliche Ressourcen (5 Items), materielle Ressourcen (5 Items), interne Kooperation (2 Items), externe Kooperation (6 Items), organisationale Stressoren (5 Items), patientenbezogene Stressoren (9 Items), Zeitdruck bei unspezifischen Festlegungen (2 Items), Zeitdruck bei spezifischen Festlegungen (6 Items), widersprüchliche Aufträge (5 Items), informatorische Erschwerungen (5 Items), Arbeitsunterbrechungen (4 Items), Qualitätseinbußen (3 Items), ungünstige Arbeitsumgebung am Einsatzort (9 Items) und ungünstige Arbeitsumgebung im RTW (4 Items) zu erfassen. Mittels eines 5-stufigen Antwortformates (1=„nein gar nicht“, 5=„ja, genau“) sollten die Teilnehmer angeben, wie sehr sie den jeweiligen Aussagen zu ihrer Tätigkeit zustimmen. Zusätzlich wurden aus dem Copenhagen Psychosocial Questionnaire, (COPSOQ; Nübling, Stöbel, Hasselhorn, Michaelis, & Hofmann, 2005) die Skalen Bedeutung der Arbeit (3 Items), Rollenklarheit (4 Items), Führungsqualität (5 Items), Unsicherheit des Arbeitsplatzes (4 Items), Emotionale Anforderungen (3 Items), Emotionale Belastungen (1 Item), soziale Unterstützung durch Vorgesetzten und Kollegen (4 Items) sowie Work – (family) privacy conflict (2 Items) eingesetzt. Emotionale Anforderungen und Belastungen wurden zum Zeitpunkt T2 erneut erfragt. Als sehr spezifischer Aspekt, der sich negativ auf das Umsetzen von im Training gelernten auswirken könnte, wurde ein nicht-straftender Umgang mit Fehlern in der Organisation (3 Items, z.B. „Die Mitarbeitenden haben das Gefühl, dass ihnen ihre Fehler zur Last gelegt werden.“) aus dem Hospital Survey on Patient Safety Culture for Hospital Management von Hammer et al. (2011) erfasst.

Als potenziell relevanter Einflussfaktor wurde Erfahrung sowohl im Bereich des Rettungsdiensts als auch im Bereich des Autofahrens wie auch einige andere Kontrollvariablen in den demographischen Daten erfasst, die alle zum Zeitpunkt T1 und auszugsweise auch zu T2 erhoben wurden. Insgesamt wurden 27 neu entwickelte Items verwendet, die zum Teil geschlossene Antwortvorgaben hatten, zum Teil aber auch freie Antwortmöglichkeiten vorsahen. Folgende Daten wurden erfasst:

- Alter
- Geschlecht
- Qualifikation im Rettungsdienst
- Hilfsorganisation
- Rettungsdienstzugehörigkeit und Hilfsorganisationszugehörigkeit
- Angestelltenverhältnis
- Vertragliche und tatsächliche Arbeitszeit
- Stunden pro Monat als Fahrer/Beifahrer
- Einsatzfahrten pro Monat
- Führerscheinbesitz: Anzahl und Dauer
- Fahrgastbeförderungserlaubnis
- Jährliche Fahrleistung
- Einträge im Fahreignungsregister („Punkte in Flensburg“)
- Unfallerrfahrung (Art und Häufigkeit) innerhalb und außerhalb des Rettungsdienstes
- Fahrsicherheitstrainings und verkehrsrelevante Fortbildungen
- Blaulichtbelehrungen
- Gesundheitsmaße: Raucherstatus, Herz-Kreislauf-Erkrankungen, Medikamente, sportliche Aktivität, Größe, Gewicht

Zur Kontrolle verschiedener Einflüsse auf die Fahrprofile der Teilnehmer (vgl. Abschnitt 0) wurden auch Daten zu den Einsatzfahrten im Rahmen der Schichterhebungen erfragt. Diese werden in Abschnitt 3.2.4 näher beschrieben.

### 3.2.2 Wissenstest: Entwicklung und Auswertung

Auf der zweiten Evaluationsebene des Lernens wurde neben den schon berichteten Einstellungsfragebögen, ein eigens entwickelter Test mit zwei parallelen Formen zur Erfassung des Wissenszuwachses, der durch das Training herbeigeführt werden kann, durchgeführt. Für die Entwicklung dieses Wissenstests wurden zunächst in Anlehnung an die Trainingsinhalte Fragen entworfen und an Personen getestet, die selbst Rettungsdienst Erfahrung oder zumindest Erfahrung in Sondersignalfahrten hatten. Die erste Fassung des Tests enthielt insgesamt 26 Fragen und wurde von zehn Personen zwischen 24 und 56 Jahren (mittleres Alter 33.1 Jahre) pilotiert. Zwei Fragen konnten von niemandem korrekt beantwortet werden, alle anderen hatten sehr divergierende Itemschwierigkeiten zwischen 0.1 und 0.91 (Mittelwert 0.43). Offene Rückmeldungen zu den einzelnen Fragen wurden genutzt, um diese verständlicher zu formulieren.

Basierend auf diesen Ergebnissen wurden aus den 26 Fragen im zweiten Schritt zwei parallele Fragebogenformen mit jeweils 20 Fragen erstellt. Dazu konnten einige Items direkt beiden Bögen zugeordnet werden. Für andere wurden parallele neue Versionen der Fragen erstellt. Teilweise wurde auch nur die Reihenfolge der Antworten verändert. Diese Versionen wurden wieder jeweils mit fünf Personen getestet, um sie auf Parallelität untersuchen zu können. Diese Personen hatten keinen Rettungsdiensthintergrund, sondern stammten aus dem Forschungsbereich. Mit dem Mann-Whitney-Test fanden sich signifikante Unterschiede für vier parallele Items, die erneut mithilfe der Anmerkungen der Tester verändert wurden, und als finale Fassungen des Wissenstests in der Studie verwendet wurden.

Die eingesetzten Wissenstests-Versionen bestehen aus jeweils 20 Fragen, wovon die meisten Multiple-Choice-Fragen mit fünf Antwortmöglichkeiten und einer richtigen Lösung sind. Zwei Fragen haben mehrere richtige Antworten, in einer Frage sollen vier Bereiche in die richtige Reihenfolge gebracht werden und eine Frage hat ein freies Antwortformat, in dem die Probanden bis zu fünf Antwortmöglichkeiten geben können. Insgesamt kann ein Punkt pro Frage erreicht werden. Bei mehreren richtigen Antworten werden die Antworten anteilig mit jeweils 0.5 bzw. 0.2 Punkten gezählt, um insgesamt auf einen Punkt zu kommen. Die freien Antworten wurden im Nachhinein mittels eines Kategoriensystems bewertet.

Die offene Frage (für beide Versionen identisch) besteht darin, dass die Probanden bis zu fünf konkrete Maßnahmen benennen sollen, um sicher am Einsatzort anzukommen. Eingeleitet wird die Frage mit einem kurzen Text, der eine bis dahin stressige Schicht mit einigen Einsatzfahrten und nun mit einem unklaren Einsatzstichwort eine Fahrt im Feierabendverkehr beschreibt. Bei Nachfrage während der Testdurchführung wurde den Teilnehmern erklärt, dass das Beispiel nur dazu dient, dass sie sich in die Situation hineinversetzen, dass sie aber auch an andere Maßnahmen denken sollen, die sie sicher zum Einsatzort bringen.

Die Bewertung der freien Antworten erfolgte mithilfe eines erstellten Kategoriensystems, das sich an den Inhalten des Trainings orientierte. So konnten insgesamt sieben Kategorien definiert werden, die auch explizit im Training behandelt werden, um Handlungsalternativen herauszuarbeiten. Zudem wurde eine weitere Kategorie aufgenommen, die sich mit allgemein vorbeugenden Maßnahmen beschäftigt. Tabelle 5 zeigt eine Übersicht der acht Kategorien.

Zur Beurteilung wurden jeweils die Antworten pro Proband in die Kategorien sortiert oder als falsch gewertet, wenn sie nicht eindeutig genug waren. Pro Kategorie konnte genau ein Punkt erreicht werden, um „Assoziationskette“ aus dem immer gleichen Bereich nicht über zu bewerten. Somit war es prinzipiell möglich, maximal 7,5 Punkte bekommen zu können, wenn alle Kategorien getroffen wurden (was in der gesamten Stichprobe aber nicht vorkam). Während der Entwicklung und Testung des Kategoriensystems wurden falsche oder zu unkonkrete Maßnahmen, für die keine Punkte vergeben werden sollten, aufgelistet und definiert, um die endgültige Bewertung möglichst objektiv durchführen zu können.

Tabelle 5: Kategoriensystem zur Beurteilung der freien Antworten im Wissenstest

Kategorie	Kategorie und Erklärung	Beispiele	Punkte
Ablenkung minimieren bzw. Prioritäten setzen	Während der Fahrt gilt es, sich auf die Fahraufgabe zu konzentrieren. In diese Kategorie fallen alle Maßnahmen, die helfen, während der Fahrt nicht durch andere Aufgaben abgelenkt zu werden	Navigation vor der Fahrt oder durch Beifahrer einstellen (lassen) Andere Störfaktoren ausschalten/minimieren (Radio, Fragen des Beifahrers, Handy, Piepser)	1
Routenplanung	Eine sinnvolle Auswahl der Route herausuchen, schon vor der Fahrt oder während, wenn man merkt, sie funktioniert so nicht	Vorher sinnvolle Strecke überlegen Busspuren/Taxispuren/Bahnspuren nutzen Ortskenntnis, um auch spontan Strecke zu ändern	1
Nutzung des Sonder- und Wegerechts	Die Entscheidung, ob es genutzt wird und wie es umgesetzt wird, fällt in diese Kategorie: Blaulicht allein gibt kein Wegerecht; es gibt Fälle, in denen sich das SoSi nicht verhältnismäßig ist	Durchgehend mit Sondersignal bzw. Einsatzhorn fahren Entscheiden, ob SoSi gebraucht wird trotz SoSi nicht davon ausgehen, dass alle richtig reagieren	1
Geschwindigkeit/Fahrphysik	Die Geschwindigkeit hat den größten Einfluss auf den Anhalteweg und die Aufprallgeschwindigkeit und entsprechend mögliche Unfallfolgen. Daher ist die Reduzierung und Anpassung an bestimmte Verkehrssituationen unerlässlich	Geschwindigkeit reduzieren Abstand vergrößern bzw. halten Defensive Fahrweise CAVE (da nicht konkret genug): Vorausschauend fahren Sicherheit vor Schnelligkeit Ruhe bewahren	1
Handlungsspielraum für andere Verkehrsteilnehmer schaffen	Anderen Fahrzeugführern wird es erleichtert sich auf ein SoSi-Fahrzeug einzustellen, wenn ihnen eindeutig angezeigt wird, wo man hin möchte und ihnen Zeit gibt eine sichere Ausweichmöglichkeit zu finden	Frühzeitige Richtungsanzeige Nicht drängeln Keine riskanten Überholmanöver Eindeutig fahren warten, wenn sich nichts mehr bewegt	1
Entscheidungsverhalten	Entscheidungsverhalten hängt mit emotionalen/motivationalen Einflüssen zusammen. Handlungskontrolle soll beibehalten werden. Subjektives Risiko erscheint geringer, bei vergrößertem Anreiz.	Einsatzstichwort/Meldebild nicht in die Fahrt einfließen lassen Emotionen nicht in Einsatzfahrt fließen lassen nicht über bevorstehenden Einsatz nachdenken	1
Informationsverarbeitung und persönliche Voraussetzungen	Nur eine gewisse Anzahl an Informationen können pro Zeiteinheit aufgenommen werden. Stress und Einsatzgeschehen können sich negativ auf diese Anzahl auswirken.	Fahrerwechsel bei Müdigkeit erwägen/durchführen Beifahrer soll aktiv mitfahren und vor Gefahren warnen Akzeptanz von Problemen/Verzögerungen	1
Vorbeugende/allgemeine Maßnahmen:	Maßnahmen zur Sicherheit, die bei jeder Fahrt (auch ohne SoSi) durchgeführt werden können, bzw. die Technik und Sicherheit des Fahrzeugs betreffen	Anschnallen Sitz/Spiegel einstellen Licht einschalten Bremsbereit sein mit Fahrzeug vertraut machen	0.5

Um zu überprüfen, ob die freien Antworten trennscharf und eindeutig von verschiedenen Beobachtern in die Kategorien eingeordnet und bewertet werden, wurden Antworten von 20 zufällig gezogenen Probanden unabhängig von zwei Ratern bewertet (da ein Proband doppelt gezogen wurde, wurden 19 Probanden genutzt) und prozentuale Übereinstimmungen sowie Intraklassen-Korrelationskoeffizienten überprüft. 85.2% der 304 Bewertungen waren dabei übereinstimmend, der ICC über beide Rater betrug  $r_2=.753$ . Da das Ergebnis nicht ganz zufriedenstellend war, wurden alle Abweichungen besprochen, die Kategorien- definitionen geschärft und die Beispiele für die jeweiligen Kategorien angepasst: Nach dieser Überarbeitung wurde ein zweiter Durchlauf: Erneut wurden 20 zufällige Probanden aus der übrigen Stichprobe gezogen und von denselben beiden Ratern bewertet. Die prozentuale Übereinstimmung der Bewertungen lag bei 93.8%. Mit einem ICC von  $r_2=.909$  ist die Beurteilerübereinstimmung sehr zufriedenstellend. Die finale Auswertung wurde von einem Rater vorgenommen, uneindeutige Antworten wurden konsensual entschieden.

Die Auswertung des Wissenstests erfolgte neben dem Gesamtsummenwert auch mit Unterpunkten zu verschiedenen Wissensbereichen, um analysieren zu können, ob bestimmte Inhalte im Training besser rezipiert werden als andere und um ggf. Handlungsempfehlungen für das Training ableiten zu können (z.B. müsste ein Inhalt stärker/anders bearbeitet werden). Die Bereiche orientieren sich ebenso wie die Kategorien der freien Antworten an den Inhalten des Trainings. Tabelle 6 zeigt die Wissenstestbereiche sowie ein Beispiel und die mögliche maximale Punktzahl, die in jedem Bereich erreicht werden kann.

Tabelle 6: Wissenstestbereiche mit Beispielitems

Bereich	Beispielitem	Mögliche Punktzahl
Unfall- ursachen und - situationen	Im Vergleich zu Normalfahrten treten Unfälle bei Sondersignalfahrten wie viel häufiger auf? <input type="checkbox"/> mit Sachschaden 10-fach häufiger auf <input type="checkbox"/> mit Schwerverletzten 8-fach häufiger auf <input type="checkbox"/> mit Todesfolge 2-fach häufiger auf <input type="checkbox"/> mit Schwerverletzten 17-fach häufiger auf <input type="checkbox"/> mit Todesfolge 7-fach häufiger auf	7
Rechtliche Grundlagen	Wie müssen Fahrten unter „Wegerechten“ (§38 der StVO) anderen Verkehrsteilnehmern deutlich gemacht werden? [nur eine Antwort] <input type="checkbox"/> gar nicht <input type="checkbox"/> durch blaues Blinklicht <input type="checkbox"/> durch blaues Blinklicht und das Martinshorn <input type="checkbox"/> durch das Martinshorn <input type="checkbox"/> durch blaues Blinklicht und Warnblinklicht	5
Fahrphysik	Sie fahren mit 30 km/h auf einer Straße. Plötzlich taucht ein Hindernis auf, Sie machen eine Vollbremsung und kommen gerade so vor dem Hindernis zum Stehen. Wie hoch wäre die Aufprallgeschwindigkeit am Hindernis, wenn Sie zu Beginn 50 km/h gefahren wären? [nur eine Antwort] <input type="checkbox"/> 10 km/h <input type="checkbox"/> 20 km/h <input type="checkbox"/> 30 km/h <input type="checkbox"/> 40 km/h <input type="checkbox"/> 50 km/h	6
Informations- verarbeitung	Wie breit ist das zentrale Blickfeld, in dem Menschen scharf sehen können? [nur eine Antwort] <input type="checkbox"/> 28° <input type="checkbox"/> 20° <input type="checkbox"/> 15°	3

Bereich	Beispielitem	Mögliche Punktzahl
	<input type="checkbox"/> 7° <input type="checkbox"/> 2°	
Entscheidungsverhalten:	Welche der folgenden Aussagen können zu einem erhöhten Verkehrsrisiko führen? [Mehrfachantworten möglich] <input type="checkbox"/> Ich finde Verkehrssicherheit sehr wichtig. <input type="checkbox"/> Ich will um jeden Preis Menschenleben retten. <input type="checkbox"/> Ich fühle mich beim Fahren eines RTW sicher. <input type="checkbox"/> Ich fahre gern schnell. <input type="checkbox"/> Ich finde die meisten Geschwindigkeitsbeschränkungen nachvollziehbar. <input type="checkbox"/> Ich finde es spannend, RTW zu fahren. <input type="checkbox"/> Ich empfinde Respekt, wenn ich mit Sonder- und Wegerechten fahre. <input type="checkbox"/> Ich finde, dass das Einsatzstichwort einen Einfluss auf die Fahrweise haben sollte.	3
Sonstige	Wie hoch ist das Unfallrisiko für einen Fahrer mit 0,6 Promille Blutalkohol im Vergleich zu einem Fahrer mit 0,0 Promille. [nur eine Antwort] <input type="checkbox"/> 1,5-fach erhöhtes Risiko <input type="checkbox"/> 2-fach erhöhtes Risiko <input type="checkbox"/> 2,5-fach erhöhtes Risiko <input type="checkbox"/> 3-fach erhöhtes Risiko <input type="checkbox"/> 3,5-fach erhöhtes Risiko	2,5

Die zwei parallelen Formen des Tests wurden möglichst gleichmäßig auf die Probanden aufgeteilt, so dass jeder Test in jeder Gruppe etwa gleich häufig als Vor- bzw. Nachtest durchgeführt wurde. Dies wurde angelehnt an eine Parallelisierung realisiert - es wurden jeweils zwei Probanden innerhalb einer Gruppe Test A bzw. B zugeordnet. Dabei wurde darauf geachtet, möglichst zwei gleichgeschlechtliche Personen aus einer Organisation zuzuteilen. So sollten in den beiden Reihenfolgeversionen jeweils vergleichbare Probanden hinsichtlich der Organisation sowie des Geschlechts sein und die Anzahl der beiden Reihenfolgen vergleichbar auf die Untersuchungsgruppen und proportional zueinander zugeordnet sind. Die folgende Tabelle gibt einen Überblick über die abschließende Verteilung der Wissenstestversionen innerhalb der Versuchsgruppen inklusive der Geschlechtsverteilung. Es gibt keinen signifikanten Unterschied zwischen den Gruppen ( $p=.964$ ) und somit können die Auswertungen im Folgenden ohne gesonderte Kontrolle der Testversionen erfolgen.

Tabelle 7: Verteilung der Wissenstestversionen innerhalb der Versuchsgruppen

Versuchsgruppe	Geschlecht	Reihenfolge		Signifikanztest
		AB	BA	Signivikanzniveau (p)
Gruppe 1 (N=71)	weiblich	7	7	.860
	Männlich	27	30	
	Summe	34	37	
Gruppe 2 (N=74)	weiblich	9	8	.782
	Männlich	28	29	
	Summe	37	37	
Gruppe 3 (N=25)	weiblich	4	3	.568
	Männlich	8	10	
	Summe	12	13	

Da der Aufwand zur Erstellung weiterer Parallelförmigkeiten den Rahmen des Projekts gesprengt hätte, wurde das Wissen - abweichend von den weiteren Evaluationsparametern – ausschließlich zum ersten und zweiten Messzeitpunkt getestet, um den Effekt des Trainings absichern zu können. Von einer längerfristigen Betrachtung wurde jedoch abgesehen.

### 3.2.3 Fahrprofile und Videodaten

Neben kurzen fragebogengestützten Erfassungen von Schichtvariablen (vgl. Abschnitt 3.2.1.3) wurden die Parameter der dritten Evaluationsebene ausschließlich objektiv mittels Fahrprofilen und Videos erfasst. Die Fahrprofile dienen dazu herauszufinden, ob die Probanden sich nach dem Training tatsächlich anders verhalten, d.h. in realen Sondersignalfahrten ihre Fahrweise ändern.

Zur Aufzeichnung der Fahrprofile wurde durch die Verkehrsunfallforschung GmbH eine App erstellt, die eine Reihe unterschiedlicher Daten während der Fahrten aufzeichnete (siehe unten). Technisch wurde die App auf Smartphones (ohne SIM-Karten) installiert, die während der gemessenen Schichten mittels Halterung an der Frontscheibe der Fahrzeuge angebracht wurden. Die App lief durchgehend und durfte nicht vom Probanden ausgeschaltet werden. Alle Daten wurden in 100Hz aufgenommen. In der ersten Version der App gab es einen Start/Stop-Button, der nicht gebraucht wurde, aber dennoch häufig gedrückt wurde, was zu vielen fehlenden Werten („Missings“) im Fahrprofil führte, da die Größe der Dateien über den gesamten Schichtzeitraum nicht verarbeitet werden konnte. Somit wurde die zweite Version ohne diesen Button konfiguriert und ab der zweiten Erhebungswelle genutzt. Zu Beginn jeder Schicht wurde die App bei stillstehendem Fahrzeug eingestellt.

Die Datenaufzeichnung begann bei einer Geschwindigkeit von 10 km/h und stoppte, wenn mindestens 50 Sekunden keine Geschwindigkeiten über 7 km/h erreicht wurden. Somit lag meist eine Messdatei pro Fahrt vor. Wenn das Fahrzeug jedoch länger im Stillstand war, wegen Stau, Bahnübergängen oder anderen Behinderungen, konnte es auch vorkommen, dass Fahrten auf mehrere Dateien verteilt waren.

Zu den gemessenen Daten gehören die Quer-, Längs- und Hochbeschleunigung des Fahrzeugs sowie Gier- (Rotation um die Hochachse), Roll- (Rotation um die Längsachse) und Nickrotationsdaten (Rotation um die Querachse). Weiterhin wurden die GPS-Daten sowie die daraus ermittelte Geschwindigkeit erfasst. Mittels Butterworth-Filter wurden die Daten so bereinigt, dass die gemessenen Daten die Beschleunigungen des Fahrzeuges angeben, egal wie das Handy im Fahrzeug angebracht wurde. Um zu viel Rauschen in den Daten zu entfernen, ohne relevante Daten zu verlieren, wurden - basierend auf den Erfahrungen der Verkehrsunfallforschung GmbH - die Daten auf 20Hz herunter gerechnet.

Zur Ermittlung einer Reihe von Kontrolldaten wurden die gemessenen Fahrten zusätzlich mittels einer an der Frontscheibe angebrachten Kamera aufgezeichnet. Genutzt wurde zunächst eine Full HD Autokamera. Mit einer Auflösung von 1080FHD 1920x1080 und ohne Aufzeichnung von Audiodaten wurden die Fahrten aufgenommen. Die erste Generation der Kamera zeichnete bis zu 25 Minuten lange Videos auf und lief in Dauerschleife. Die Aufzeichnung begann, sobald sie über die Zündung Strom bekam, und hörte 10 Sekunden nach Ausschalten der Zündung auf. Leider konnten aufgrund einer Reihe von Defekten und Problemen sowie späterer Nicht-Mehr-Verfügbarkeit nicht während der gesamten Erhebungen die gleichen Geräte genutzt werden und es kam mit weiteren Modellen zu jeweils etwas anders gelagerten Schwierigkeiten und Problemen: Art der Stromversorgung und damit einhergehende notwendige Eingriffe durch die Versuchspersonen, aber auch die Dauer der maximal aufgezeichneten Videos inklusive der Einstellung zum Überschreiben alter Daten, wenn der Speicher (SD-Karten) voll war, waren den Herstellerangaben oft nicht zu entnehmen und variierten zwischen den Geräten stark. Diese Umstände führten leider zum Teil zu erheblichen Verlusten von Daten. Entsprechend den Vorgaben des Ethikvotums zum Datenschutz wurden die SD-Karten und Abschlüsse der Kamera vor jeder Schicht versiegelt, um ein Auslesen der Daten und Nutzung durch Externe zu verhindern.

Neben der Erfassung von Kontrolldaten dienten die Videodaten auch dazu, die Uhrzeit und Dauer der Einsatzfahrten zu bestimmen sowie die Fahrprofilaten der erfassten Schichten automatisiert in verschiedene Fahrttypen aufzusplitten (die für die bayerischen Einsatzfahrten vorliegenden Statusmeldungen sind nicht genau genug für diese Sortierung). Mit der Nutzung der Videodaten konnten Beginn und Ende der Einsatzfahrten festgelegt werden, wobei die Dauer als Proxy für die Hilfsfrist in den Auswertungen verwendet wird, und „normal ablaufende“ Fahrten zwischen Wache, Einsatzort, Klinik und Wache konnten relativ einfach identifiziert und sortiert werden. Aufwändiger war die Aufbereitung von Fahrten, in denen die Besatzung während einer Nicht-Einsatzfahrt alarmiert wurde oder eine Einsatzfahrt auf der Strecke abgebrochen wurde und somit in einer gemessenen Fahrprofildatei sowohl eine Einsatzfahrt als auch eine Nicht-Einsatzfahrt steckte. Auch Patiententransporte, bei denen erst unterwegs entschieden wurde, doch mit Sondersignalen weiter zu fahren, mussten händisch vorbereitet und gesondert aufgeteilt werden. Zur weiteren Auswertung und Nutzung der Videodaten vergleiche Abschnitt 7.3.

Zur Automatisierung der Sortierung der Einsatzfahrten wurde ein Programm geschrieben, das basierend auf einer Excel-Datei mit allen relevanten Daten (vgl. auch Einsatzinformationstabelle, Abschnitt 3.2.4) alle Fahrten der Probanden in drei verschiedene Dateien sortierte: Einsatzfahrten mit Sondersignal, Einsatzfahrten ohne Sondersignal und restliche Fahrten. In diesen Dateien wurde auch automatisch kodiert, um welchen Messblock (T1-T4) es sich bei der Fahrt handelt, ob ein RTW oder NEF gefahren wurde, ob ein Patient transportiert wurde, ob es eine Hinfahrt oder ein Patiententransport war und ob in der Aufzeichnungsschicht beim Probanden gleichzeitig ein EKG gemessen wurde. Bevor die Daten sortiert wurden, wurde ermittelt, ob Video- und Handydaten die gleiche Uhrzeit aufwiesen. Die Geräte wurden immer, bevor sie zum Probanden gebracht wurden, synchronisiert, so dass sie genau die gleiche Uhrzeit hatten. Teilweise lagen die Geräte aber einige Tage in der Wache, da die Probanden mehrere Schichten selbständig messen sollten. Daher kam es vor, dass die Geräte einige Sekunden bis Minuten Verschiebungen aufwiesen. Um diese Fehler nicht systematisch in die Sortierung einfließen zu lassen, wurde zunächst ermittelt, ob eine Verschiebung vorliegt und bei der Sortierung wurde sie berücksichtigt. Nachdem alle Daten sortiert wurden, wurde eine abschließende Kontrolle durchgeführt und händisch jeweils das Ende der Fahrprofilaten entfernt. Dies konnte nicht automatisiert erfolgen, da die Aufzeichnungen nicht immer genau 50 Sekunden länger waren als die Fahrt (siehe oben). Daher wurden nur diejenigen Daten entfernt, in denen das Fahrzeug quasi stand (weniger als 1 km/h). Da das evaluierte Training auf Sondersignalfahrten abzielt, werden in den Analysen nur diese Fahrten (v.a. der Normalfall der Fahrten zum Einsatzort sowie der seltenere Fall von Patiententransporten) betrachtet.

Zur Analyse der Daten wurden aus den sortierten Fahrprofilen eine Reihe von Kennwerten erstellt. Um eine Baseline aller Daten zu haben, wurde zunächst die Dauer der einzelnen Fahrten bestimmt. Für jeden Messzeitpunkt (T1 bis T4) werden die Werte nach Art der Fahrt (Hinfahrt und Patiententransport) und Fahrzeug (RTW und NEF) aufgeteilt, so dass insgesamt jeweils 16 Einzelwerte pro Kennwert vorliegen. Es wurde die Geschwindigkeit (Minimum, Maximum, Mittelwert, Median, 25. Und 75. Perzentil) berechnet sowie jeweils das Maximum, der Mittelwert und der Median der positiven und negativen Längs-, Quer- und Hochbeschleunigung sowie der drei Drehachsen. Zusätzlich wurden Trigger berechnet, die angeben sollen, wie oft die Probanden außerhalb eines bestimmten Bereiches fahren. Als Grundlage der Trigger wurden die Beschleunigungen der 100 car naturalistic driving study von Dingus et al. (2006) herangezogen, die auf „riskante“ respektive sehr starke Fahrbewegungen hinweisen. Eine hohe Anzahl von Triggern kann dabei als ein Maß für erratisches Fahren verstanden werden. Für die Längsbeschleunigung werden Werte über 0,6 g bzw. 5,89 m/s<sup>2</sup> als Trigger genommen. Für die Querbeschleunigung werden Werte über 0,7 g bzw. 6,87 m/s<sup>2</sup> genutzt. Da diese Werte alle auf PKW-Messungen beruhen, sind sie für Rettungswagen relativ streng - wenn Probanden diese Bereiche auch mit einem großen RTW erreichen, fahren sie vermutlich bereits recht nahe an den Grenzbereichen des Fahrzeugs. Um einen Trigger, der eher RTW angemessen ist, zu finden, wurde zusätzlich ein Fahrzeug des Rettungsdienstes während eines Fahrsicherheitstrainings mithilfe der Studiengeräte gemessen und anschließend analysiert, wie stark Vollbremsungen sich in der Längsbe-



schleunigung zeigen. Dabei wurden 17 Vollbremsungen aus 30-70 km/h analysiert. Der Untergrund (trocken und nass) zeigte keinen Einfluss auf die Höhe der Bremsbeschleunigung. Nach Ausschluss der zögerlichen Bremsvorgänge konnte ein neuer Trigger für Vollbremsungen ermittelt werden, der bei  $3 \text{ m/s}^2$  lag. Maximale Werte lagen bei  $7,7 \text{ m/s}^2$  und im Schnitt bremsen die Fahrer im Fahrsicherheitstraining mit  $5,6 \text{ m/s}^2$ .

Somit ergeben sich insgesamt zwei Trigger für Vollbremsungen ( $-3 \text{ m/s}^2$  und  $-5,89 \text{ m/s}^2$ ), ein Trigger für positive Längsbeschleunigungen ( $5,89 \text{ m/s}^2$ ) und ein Trigger für Querschleunigungen (Zusammenfassung von  $-6,87 \text{ m/s}^2$  und  $6,87 \text{ m/s}^2$ ). Es wurde ermittelt, wie viele Zeiteinheiten die Probanden insgesamt innerhalb dieser kritischen Bereiche gefahren sind, wie oft sie insgesamt in diese Bereiche gekommen sind und wie lange sie demzufolge im Schnitt in kritischen Bereichen gefahren sind. Diese Werte werden jeweils an der Gesamtdauer, die gefahren wurde, relativiert. Somit hätte ein Proband, der in 10 Minuten zweimal in einen kritischen Bereich kommt, den gleichen Wert wie ein Proband, der in 20 Minuten viermal in diesem Bereich fährt.

### 3.2.4 Schichtbegleitmessungen: Einsatzdaten, EKGs und Beanspruchung während der Schicht

Während der Schichten, in denen auch Fahrprofile und Videodaten erfasst wurden, wurden auch Beanspruchungs- und Einsatzdaten mittels Fragebogen erfasst, die vor, während und nach der Schicht durch die Probanden aufgefüllt werden sollten, sowie in einzelnen Schichten EKGs abgeleitet.

Vor der Schicht sollten zunächst einige organisatorische Daten der Schicht angegeben werden: das Datum, die Nummer der Messgeräte, Schichtbeginn und -ende, das Leitstellengebiet und der Fahrzeugtyp. Danach wurde die Basler Befindlichkeitsskala (BBS; Hobi, 1985) und die Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990) gemessen (zur Beschreibung dieser Skalen vgl. Abschnitt 3.2.1.4). Diese Daten sollten direkt vor der Schicht - vor allem vor dem ersten Einsatz - ausgefüllt werden.

Danach folgte eine Tabelle, in die alle Einsätze eingetragen werden sollten mit der Einsatznummer, der Alarmzeit, der Art des Einsatzes, der Sondersignalnutzung auf dem Weg zum Einsatzort, dem Einsatzstichwort sowie für den Patiententransport, ob selbst transportiert wurde, wann der Transport begann, der Verdachtsdiagnose und der Sondersignalnutzung. Ob der Proband selbst den Patienten transportiert hat ist wichtig, um Fahrten, die von jemand anderem durchgeführt wurden aus den Daten entfernen zu können. So konnte es vorkommen, dass der Proband eigentlich die medizinisch höhere Ausbildung hatte und somit den Patienten versorgen muss, aber um genug Messungen durchführen zu können, auch die Hinfahrten zum Einsatzort als Fahrer durchgeführt hat. Diese Einsatzinformationstabelle konnte während der Schicht oder am Ende ausgefüllt werden und auch der Beifahrer konnte sie ausfüllen, wenn der Proband beispielsweise als Fahrer auf dem Rückweg zur Wache war.

Zusätzlich während der Schicht sollten jeweils zu einem Viertel der Schicht die sechs Skalen des NASA-Task Load Index (Hart & Staveland, 1988; vgl. Abschnitt 3.2.1.4) sowie die KSS ausgefüllt werden. Die Teilnehmer waren angehalten, diese möglichst zu der Zeit auszufüllen, die vorgegeben war (also beispielsweise alle 2 Stunden bei einer 8-Stunden-Schicht). Die Zeit des Ausfüllens sollte zur Kontrolle zusätzlich angegeben werden, um sicherzustellen, dass die Abstände zwischen den Messungen hinreichend weit auseinander lagen. Zu Schichtende sollte noch einmal die Basler Befindlichkeitsskala, der NASA sowie die KSS ausgefüllt werden. Zusätzlich sollten die Probanden angeben, wie sie ihre eigene Fahrleistung während der Schicht bewerten würden und ob sie eine kritische Situation während einer Einsatzfahrt erlebt hatten. Auch die Beifahrer sollten (sofern die Fahrer in diese Beifahrerbefragung eingewilligt hatten) die Einsatzfahrten bewerten. Aufgrund des normalen Rettungsdienstalltags waren in diesen schichtbegleitenden Messungen die meisten Lücken zu verzeichnen - viele Bögen wurden aufgrund von Einsätzen oder auch Schlaf während

der Nachtschichten nicht zu den geplanten Zeiten, nur teilweise oder auch gar nicht ausgefüllt. Vor diesem Hintergrund musste mit Blick auf die Beanspruchung auf eine Auswertung von Verläufen während der Schicht verzichtet werden, stattdessen werden aggregierte (Mittel- oder Deltawerte) in den Analysen verwendet.

Als objektives Maß für die Höhe der Anspannung der Einsatzfahrer wurde während einiger Schichten (jeweils eine pro Messzeitpunkt) ein Langzeit-EKG (5- respektive 7-Kanäle) gemessen. Die ursprünglich geplante Auswertung dieser EKGs mit Blick auf Herzratenvariabilität (HRV) als psychophysiologischem Link zu stressauslösenden Arbeitsbelastungen (vgl. z.B. Jarczok et al., 2013) konnte aufgrund ähnlicher Probleme wie bei den weiteren Schichtmessungen nicht durchgeführt werden – körperliche Anstrengung und Schwitzen bei den Einsätzen führten zu einem vergleichsweise hohen Anteil an abschnittswisen Fehlmessungen (bis die Probanden bemerkten, dass sich Elektroden gelockert hatten), die insbesondere bei zeitbasierten HRV Parametern zu fehlender Reliabilität führen würden. Daher wurde als ein deutlich robusteres Maß zur Auswertung die minimale, maximale und mittlere Herzfrequenz als relevanter körperlicher Parameter in den Analysen verwendet. Für diese stellte Krüger (2010) in einem Vortrag die hohe physiologische Beanspruchung während Fahrten unter Nutzung von Sonder- und Wegerechten vor und zeigte, dass die Herzfrequenz sowohl für Novizen als auch für Experten während einer Einsatzfahrt auf 140-160 Schläge pro Minute anstiegen im Vergleich zu 70-100 Schlägen pro Minute während Vergleichsfahrten.

Aus den Schichtdaten wurde auch ein wahrscheinlich hochrelevanter moderierender Faktor sowohl im Hinblick auf die Fahrprofilaten als auch im Hinblick auf die Beanspruchung codiert – die (von den Rettungsdienstmitarbeitern vermutete) Einsatzschwere: Es ist anzunehmen, dass sich der Stress und die Fahrweise je nach Einsatzstichwort unterscheidet. So sollten etwa Notfälle mit Beteiligung von Kindern oder Großschadensereignisse das Stressniveau deutlich anheben. Um diese Einsatzschwere zu erfassen, wurden bei den Hinfahrten mit Sondersignal die Einsatzstichwörter aus der Einsatzinformationstabelle und bei Patiententransporten mit Sondersignal die Verdachtsdiagnosen verwendet. Die Kodierung erfolgte pro Einsatz mit drei Schweregraden von Schwere 1 („eher unkritisch“) bis 3 („kritisch“) und in Anlehnung an den Notarztindikationskatalog der Bundesärztekammer (Stand 2013). So wurden z.B. alle Einsätze mit akuter vitaler Bedrohung (wie Reanimationen, Herzinfarkte, Schlaganfälle oder Krampfanfälle) sowie weitere des Notarztindikationskataloges mit einer Schwere von 3 kodiert. Einige Einsätze wie Verkehrsunfälle, Atemnot, Anaphylaxie und Intoxikationen wurden nur dann mit „kritisch“ kodiert, wenn auch tatsächlich ein Notarzt mit alarmiert wurde. Kindernotfälle galten als solche, wenn der Patient bis zu 12 Jahre alt war, während Angaben bei älteren Kindern und Jugendlichen mittels des weiteren Einsatzstichwortes bewertet wurden. Als „eher unkritisch“ wurden Einsätze kodiert, bei denen bereits Einsatzkräfte oder medizinisches Personal vor Ort waren und somit keine übermäßige Eile bei der Anfahrt geboten war. Hierzu zählen Krankentransporte, Verlegungen, Hilfeleistungen und Nachforderungen. Auch das (häufiger vorkommende) Stichwort „Brandmeldeanlagen“ wurde als „eher unkritisch“ kodiert, da hier kein gesicherter Brand vorliegt, oft Fehlalarme vorliegen und keine verletzten Personen vermutet werden. Alle weiteren Einsätze wurden mit der Schwere 2 („eher kritisch“) kodiert. Hierunter fallen Stichworte wie Kopfplatzwunde, verletzte/erkrankte Person, Sturz, Nasenbluten, Hyperventilation, Bluthochdruck oder Kollaps. Auch Einsätze, bei denen kein Einsatzstichwort oder Verdachtsdiagnose angegeben war, wurden mit der Schwere 2 kodiert (insgesamt 2.5% der Einsätze auf der Hinfahrt und 2.6% der Patiententransporte).

In diesem Abschnitt wurden alle Erhebungsmethoden des Projekts nach Art und Evaluationsebene mit den Angaben zur psychometrischen Güte vorgestellt:

- Neben Fragebögen werden in der Studie auch Wissenstests, Fahrprofile, Videodaten und schichtbezogene Erhebungen inkl. Schicht-EKGs verwendet.

- Für die Reaktionsebene wurde ein Fragebogen zur Trainingsevaluation mit den Skalen Inhalte, Trainer, Rahmenbedingungen und Anwendbarkeit neu entwickelt.
- Auf der Lernebene werden Fragebögen zur Erfassung verkehrsrelevanter Einstellungen und Persönlichkeitsmerkmale eingesetzt. Aufgrund schlechter psychometrischer Werte wurden neue Skalenstrukturen für den Test verkehrsrelevanter Persönlichkeitsmerkmale und für die Risikowahrnehmung im Straßenverkehr aus dem Risk Perception Questionnaire und dem Fragebogen zur Risikowahrnehmung bei Regelverstößen entwickelt. Weitere Fragebögen/Skalen (u.a. Risk Seeking, Driving Skill Questionnaire) wurden wie aus der Literatur bekannt übernommen oder völlig neu entwickelt (z.B. Spaß und Beanspruchung während Einsatzfahrten).
- Ebenfalls für die Lernebene wurde ein Wissenstest mit zwei parallelen Testformen basierend auf den Inhalten des Trainings neu entwickelt und validiert.
- Auf der Verhaltensebene werden Fahrprofilaten und Videoaufzeichnungen genutzt: Mithilfe einer Handy-App wurden Beschleunigungen und Gierwinkel jeweils in drei Achsen sowie Geschwindigkeiten während der Einsatzfahrten aufgezeichnet und Videodaten werden zur Kontrolle und Darstellung der Einsatzfahrten genutzt. Parameter zur Beschreibung der Fahrprofile (u.a. Trigger für kritisches Fahren) wurden entwickelt.
- Etablierte Fragebogenverfahren zu kurz- und längerfristigen Beanspruchungsfolgen werden zur Evaluierung der Resultateebene eingesetzt. Die kurzfristigen Folgen wurden mit Basler Befindlichkeitskala, Karolinska Sleepiness Scale und NASA-Taskload Index mehrfach während der erfassten Schichten erhoben. Zudem werden EKGs zur Beanspruchungsmessung während der Schichten und Einsatzfahrten eingesetzt.

### 3.3 Analysemethoden

Der statistische „Königsweg“, um in einem kontrollierten Design wie dem vorliegenden Veränderung und Kausalität von Veränderung zu prüfen, ist die Varianzanalyse. Die Varianz einer oder mehrerer Zielvariablen wird dabei durch den Einfluss einer oder mehrerer Einflussvariablen (Faktoren) erklärt. Prüfgrößen (F-Werte) geben an, ob die Varianz zwischen den Gruppen statistisch bedeutsam (signifikant) größer ist als die Varianz innerhalb der Gruppen. Für die hier vorliegende Evaluation ist eine zweifaktorielle Varianzanalyse mit Messwiederholung das Verfahren der Wahl. Faktor eins entspricht dabei den drei Gruppen Versuchsgruppen, Faktor zwei ist der Messwiederholungsfaktor. Da keine vollständige Messung über alle vier Zeiten vorliegt, sind die Stufen des Faktors in den verschiedenen Analysen unterschiedlich, das Hauptdesigns zur Erfassung der durch das verursachten Veränderungen umfasst den ersten und zweiten Messzeitpunkt (vgl. Abschnitt 4.2), die Frage, ob potenzielle Effekte in der Wartegruppe replizierbar sind, umfasst die Zeitpunkte T2 und T3 (vgl. Abschnitte 4.3.1, 4.4.1 und 4.5.1), und die Frage der Nachhaltigkeit wird mit einem Vergleich der Zeitpunkte T1 und T4 untersucht (vgl. Abschnitt 4.3.2, 4.4.2 und 4.5.2). Varianzanalysen dieser Form liefern jeweils Haupteffekte für die einzelnen Faktoren sowie Interaktionseffekte. Im strengen Sinn ist die Wirksamkeit einer Intervention nur dann gegeben, wenn sich eine signifikante Interaktion in die erwartete Richtung zeigt. Die aufgetretenen Interaktionen und Haupteffekte werden im Ergebnisteil jeweils erläutert.

Die Voraussetzungen zur Durchführung von Varianzanalysen (Normalverteilung der abhängigen Variablen sowie Varianzhomogenität) wurden jeweils überprüft. Da sich keine systematischen Verletzungen der Voraussetzungen fanden, werden im Folgenden jeweils die Ergebnisse zweifaktorieller Varianzanalysen mit Messwiederholung auf die Variablengruppen der verschiedenen Evaluationsebenen berichtet. Um keine Effekte zu übersehen, wird ein erhöhtes Signifikanzniveau von  $\alpha = .10$  angelegt und es werden die

Effektstärken näher betrachtet. Aufgrund einer hohen Dropoutquote auf verschiedenen Ebenen des Studiendesigns (vgl. Abschnitte 3.4.2 und 3.4.3 zur Stichprobe und Zellbesetzungen) werden neben den zu präferierenden längsschnittlichen Analysen auch Varianzanalysen durchgeführt, bei denen die Messzeitpunkte unabhängige Faktoren sind (ungepaarte Stichprobe). In diesen Analysen können keine Aussagen mehr über Veränderungen getroffen werden, sie haben jedoch den Vorteil, dass auch Personen mit fehlenden Messungen zu einem Zeitpunkt einbezogen werden können, damit die statistische Power erhöht wird und auch die Werte dieser Personen Berücksichtigung finden.

Bei allen Analysen, die nicht unmittelbar dem Studiendesign folgen, werden inferenzstatistische Testverfahren in Abhängigkeit vom Niveau und der Verteilung der Daten angewendet. Dabei wird wann immer möglich auch eine Einordnung der Befunde durch die entsprechenden Referenzwerte durchgeführt. Bei einzelnen Subgruppenvergleichen kann es aufgrund kleinerer Gruppengrößen leichter zur Verletzung der Verteilungsvoraussetzungen kommen, so dass bei diesen Analysen ggf. auf nichtparametrische Verfahren zur Auswertung zurückgegriffen wird. Nichtparametrische Verfahren, v.a. Rangreihentests, sind robust gegenüber Verletzungen der Voraussetzungen und werden häufig routinemäßig bei kleinen und sehr kleinen Stichproben eingesetzt. Spezifische weitere statistische Verfahren werden im Ergebnisteil jeweils kurz dargestellt. Alle Analysen wurden mit dem Programm-Paket IBM SPSS Statistics 24 durchgeführt.

### 3.4 Stichprobe und Zellbesetzungen

Im folgenden Abschnitt werden, abweichend von der üblichen Darstellung, neben der endgültigen Stichprobe, auch Akquise und Schwierigkeiten bei der Akquise, teilnehmende Organisationen, Zusammensetzungen der Trainings und anderes mehr berichtet, um dem Leser ein möglichst klares Bild der Teilnehmer zu vermitteln und die Notwendigkeit der aufwändigen Einzelanalysen ab Abschnitt 4.3 zu erläutern.

#### 3.4.1 Akquise

Auf Basis von Fallzahlabeschätzungen, die mit dem „härtesten“ Evaluationskriterium, dem der Fahrprofile, vorgenommen wurden, wurde von einer benötigten Anzahl von 200 Teilnehmern ausgegangen. Bei einer geschätzten Ausfall-/Dropoutquote von 20% sollten damit insgesamt 240 Personen aus zwei Großräumen (aufgrund der Simulatorverfügbarkeit und zur Erhöhung der Generalisierbarkeit der Befunde) akquiriert werden.

Geplant war eine erste Erhebungswelle in einem bayerischen Großraum und darauffolgend eine zweite in einem sächsischen Großraum, Leider zeigten sich bereits früh große Probleme bei der Teilnahmebereitschaft der Rettungsdienste, der einzelnen Mitarbeiter und bei der Compliance in der Datenerhebung, so dass zum einen im Design Untersuchungsgruppe 4 entfernt wurde (vgl. Abschnitt 3.1) und zum anderen gegen Projektende eine dritte Erhebungswelle in Bayern durchgeführt werden musste. Zudem wurde nach einer Teilnahmeabsage durch einen Rettungszweckverband musste der Schwerpunkt der zweiten Erhebungswelle auf einen anderen Großraum verschoben bzw. ausgeweitet werden. Angesprochen wurde jeweils jeder Rettungsdienstanbieter der entsprechenden Region. Tabelle 8 gibt einen Überblick über die angesprochenen Rettungsdienstorganisationen, die Anzahl der teilnehmenden Verbände pro Organisation und die Anzahl der dort vorhandenen und tatsächlich erhobenen Wachen in Bayern und Sachsen. Aus Datenschutzgründen werden die deutlich kleineren privatwirtschaftlichen Anbieter jeweils zusammen unter andere Anbieter berichtet.

Ablehnungsgründe waren insgesamt sechsmal eine Absage durch den jeweiligen Rettungszweckverband, fünfmal wurden fehlende Zeit/Kapazitäten genannt, dreimal kam es zu einer Ablehnung durch den Personal-/Betriebsrat, insgesamt fünf Verbände hatten Interesse an der Teilnahme, konnten aber keine

interessierten Mitarbeiter finden, dreimal wurde dezidiert kein Interesse angegeben, und jeweils zweimal wurden kein Grund angegeben oder die Ansprechpartner waren nicht mehr erreichbar.

Tabelle 8: Organisationsakquise aller in Frage kommenden Rettungsdienste

Rettungsdienstorganisation	Angesprochene Verbände	Teilnehmende Verbände	Geschätzte Größe*	gemessene Wachen
<b>Bayern</b>				
Arbeiter Samariter Bund	1	1	3 Wachen	2
Bayrisches Rotes Kreuz	19	10	77 Wachen	30
Johanniter Unfall Hilfe	4	2	6 Wachen	3
Malteser Hilfsdienst	2	1	5 Wachen	3
Andere Anbieter	3	3	16 Wachen	9
<b>Sachsen</b>				
Arbeiter Samariter Bund	3	2	6 Wachen	3
Deutsches Rotes Kreuz	15	5	66 Wachen	13
Johanniter Unfall Hilfe	1	0	1 Wache	0
Malteser Hilfsdienst	3	1	10 Wachen	2
Andere Anbieter	1	1	2 Wachen	2
<b>Summe:</b>	<b>52 angesprochene Verbände</b>	<b>26 Teilnahmen (50%)</b>	<b>192 Wachen</b>	<b>67 Wachen (34.9%)</b>

\*geschätzte Größe nur bezogen auf die angesprochenen Ortsverbände

### 3.4.2 Stichprobe und Dropout-Analyse

Abbildung 4 gibt einen Überblick über alle Teilnehmer und Dropouts im kompletten Erhebungsverlauf. Insgesamt bekundeten 217 Mitarbeiter der teilnehmenden Organisationen Interesse an einer Teilnahme, von denen 27 aus studientechnischen Gründen (z.B. zu geringe Fahrzeiten) aber auch aus eigenem Antrieb (z.B. weil in der entsprechenden Organisation die Studienteilnahme nicht als Arbeitszeit abgegolten wurde, oder weil eine Schwangerschaft eintrat) bereits vor Studienbeginn ausfielen. Für die folgenden relevanten Erhebungsschritte wird in der Übersicht zwischen freiwillig aus der Studie ausgestiegenen Teilnehmern (Dropout) und ausgeschlossenen Teilnehmern (Ausschluss) unterschieden. Ausgeschlossen wurden Probanden, die ihre Eingangs- bzw. Abschlussfragebögen (T1-T4-Bögen) nicht abgegeben haben. Diese sind unentbehrlich für viele Kontrollvariablen und zur Messung der Einstellungsänderung.

Ebenso wurden Probanden ausgeschlossen, bei denen ein zweiter Wissenstest fehlte resp. nicht erhoben werden konnte. Die weiteren Ausschlussgründe betrafen die Schichtmessungen und damit die dritte und vierte Evaluationsebene. Personen, die keine Handydaten oder keine Videodaten in der Ebene aufweisen können, mussten ausgeschlossen werden. Weiterhin wurden Probanden ausgeschlossen, die zu wenige Fahrprofilaten (Handy und Video zusammen) aufwiesen. Da die Fahrzeiten sehr unterschiedlich waren, wurde als Grenze ein Wert von mindestens 10 Minuten Einsatzfahrt angesetzt, um nicht zu viele Probanden ausschließen zu müssen. Als letzter Schritt mussten auch die Probanden entfernt werden, die so gut wie gar keine Schichtdaten hatten. Auch hier wurde ein geringes Minimum angesetzt, um nicht zu viele Probanden auszuschließen. Es musste mindestens ein Schichtende-Bogen einer Schichtmessung ausgefüllt sein, unabhängig, ob in der entsprechenden Schicht alle Handy- oder Videodaten vorhanden waren. Die Ausschlusskriterien waren nötig, um eine Mindestgüte für die weitere Auswertung gewährleisten zu können, da es insgesamt bei nur 574 von 1468 gemessenen Schichten keinerlei Probleme mit Handy-, Video- oder Schichtdaten (vgl. Abschnitt 0) gab.

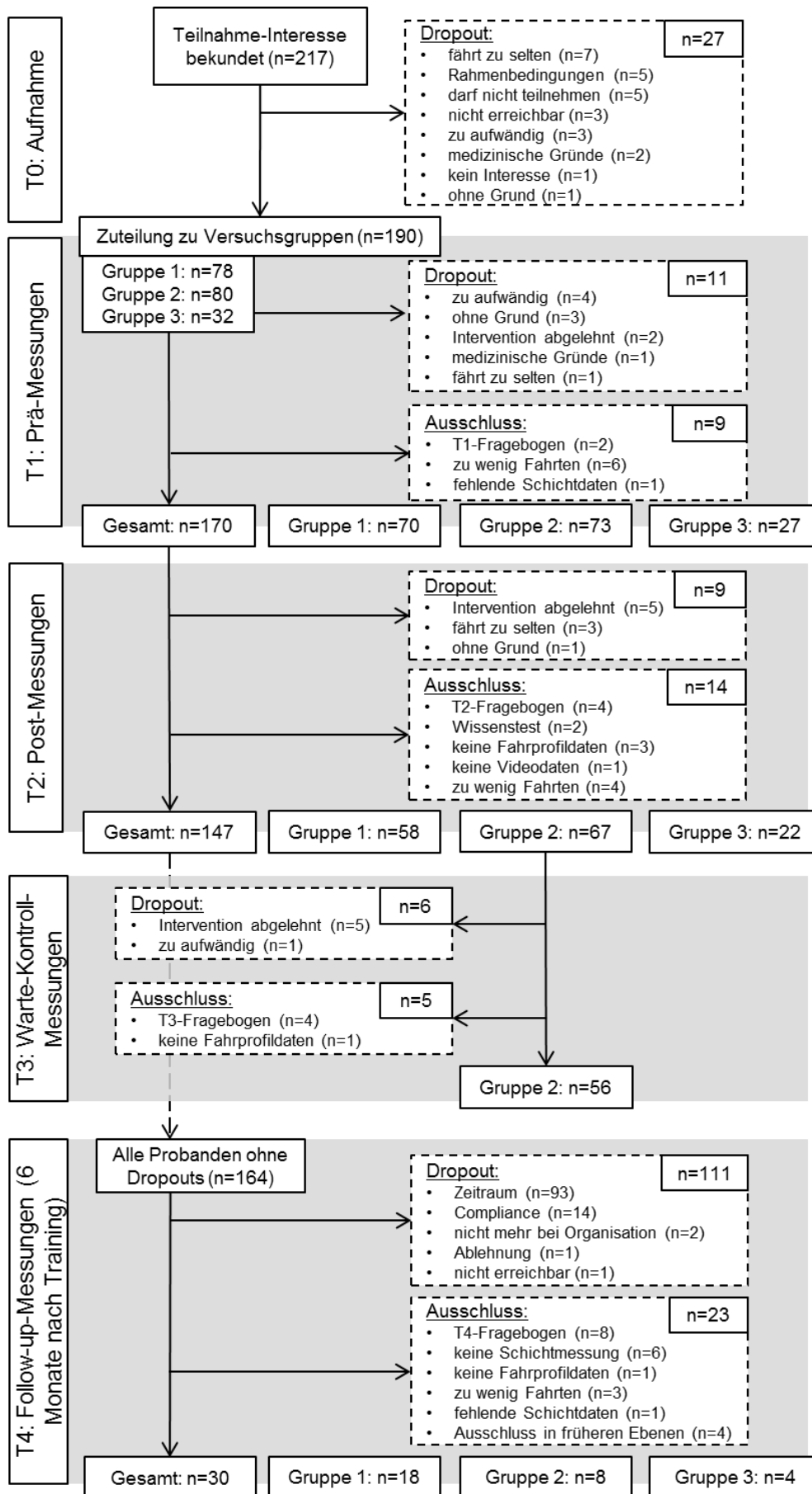


Abbildung 4: Teilnehmer und Dropout-Übersicht nach CONSORT-Schema

Im Studienablauf wurden nach der Aufnahme als Teilnahmeinteressierte, die verbleibenden 190 Probanden zu etwa gleichen Teilen in Gruppe 1 und 2 und zu einem kleineren Teil in die Gruppe 3 zugeteilt. Während der T1-Messungen traten elf Personen von der weiteren Teilnahme an der Studie zurück, größtenteils mit der Begründung, dass die Teilnahme zu aufwändig wäre. Neun Personen mussten zusätzlich aus den dargestellten Gründen ausgeschlossen werden. In der T2-Phase sind insgesamt 9 Personen freiwillig von der Studie zurückgetreten, die meisten, da sie trotz ausreichender Vorinformation das Training ablehnten und sich trotz vieler Versuche und Angebote kein „passender“ Termin fand. Erneut mussten insgesamt 14 Probanden ausgeschlossen werden.

In der Wartegruppe, die als einzigen zu T3 gemessen wurde, da für sie nach T2 das Training stattfand, nahmen ebenfalls fünf Personen nicht am Training teil, eine weitere wollte explizit nicht weiter teilnehmen. Damit war der Dropout aufgrund Verweigerung der Trainingsteilnahme mit knapp 7.5% in der Wartegruppe höher als in den Interventionsgruppen 1 und 3 mit knapp 5.2% (5 von 97). Zudem mussten fünf zusätzliche Personen aufgrund fehlender Werte ausgeschlossen werden; dieser Anteil ist mit 7.5% nur etwa halb so groß wie in den Interventionsgruppen (14.4%). Insgesamt ist die Quote für „echte“ Dropouts zu den drei Hauptmesszeitpunkten mit 13.7% (26 von 190 Teilnehmern) geringer, als ursprünglich auf Basis anderer Studien angesetzt 20%.

Die T4-Follow-up-Messungen zur Überprüfung der Nachhaltigkeit von Trainingseffekten sollten mindestens sechs Monate nach der letzten Messung stattfinden. Für die insgesamt 164 Probanden, die nicht als Dropouts die Studie verlassen hatten, wurde dieser Zeitraum überprüft. Für den größten Teil (n=93) war der Abstand zur letzten Messung zu kurz, weitere 14 wurden nicht zur Teilnahme angefragt, dass sich bei ihnen in den Ausschluss- und Kontaktdaten eine erheblich mangelnde Compliance gezeigt hatte. Dabei handelte es sich um Personen, die schon frühere T1-T3-Fragebögen nicht ausgefüllt hatten, mehrfach aufgefordert werden mussten, um ihre Messungen überhaupt durchzuführen oder sehr oft ohne Angabe von Gründen nicht an vereinbarten Terminen gemessen hatten oder so viele fehlende Daten in den Unterlagen hatten, dass die gemessenen Schichten kaum auswertbar waren. Von den verbleibenden 57 angefragten Teilnehmern stimmten 93% einer erneuten Teilnahme zu; eine Person lehnte eine weitere Teilnahme ab, eine war nicht erreichbar und zwei hatten mittlerweile die Organisation gewechselt. Trotz dieser guten Quote mussten nach abschließender Prüfung 23 Personen aufgrund weiterer fehlender Messungen oder bereits großer Probleme mit fehlenden Werten in den vorangegangenen Messungen von der Gesamtauswertung ausgeschlossen werden.

### **3.4.3 Zellbesetzungen auf den Evaluationsebenen und zu den Messzeitpunkten**

Vor dem Hintergrund des komplexen Datenmodells, das in Abbildung 5 noch einmal visualisiert ist, und der großen Menge aufgrund fehlender Werte und Messungen für eine Gesamtauswertung ausgeschlossener Personen, sollen diese Daten für Einzelauswertungen auf den verschiedenen Evaluationsebenen und zu den unterschiedlichen Messzeitpunkten nicht verloren gehen. Zu diesem Zweck zeigt die folgende Tabelle 9, wie viele Probanden in den verschiedenen Evaluationsebenen, trotz Ausschlüssen auf verschiedenen zeitlichen Ebenen, ausgewertet werden können.

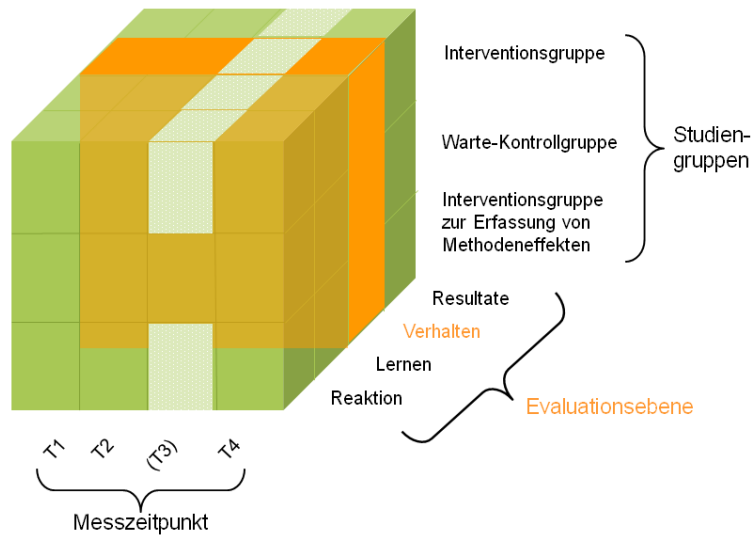


Abbildung 5: Datenmodell der Studie

Aufgeführt sind in der Tabelle auch die Anzahl der EKG-Daten. Fehlende Werte in diesem Bereich waren kein Grund für einen Ausschluss aus der Gesamtauswertung, da dieser Aspekt der Resultateebene recht weit von den Trainingseffekten entfernt ist und daher als nicht zentral für die Evaluation gesehen wird. Deutlich erkennbar ist, dass für Einzelvergleiche und Auswertungen außerhalb des vollständigen Längsschnittsdesigns erheblich mehr Daten zur Verfügung stehen.

Tabelle 9: Zellbesetzungen auf jeder Evaluations- und zeitlichen Ebene

Evaluations-ebene/ zeitliche Ebene	Reaktion	Lernen – Wissen	Lernen – Einstellung	Verhalten	Resultate – Schichtbögen	Resultate – EKGs	Gesamt-N
<b>T1</b>							
Gruppe 1:	73	76	77	75	79	69	70
Gruppe 2:	/	76	76	75	76	66	73
Gruppe 3:	25	27	25				27
<b>T2</b>							
Gruppe 1:	64	66	71	65	71	54	58
Gruppe 2:	68	74	74	70	75	62	67
Gruppe 3:	23	25	25	22	25	22	22
<b>T3</b>							
Gruppe 2:	64	64	/	57	65	47	56
<b>Gesamt-N durchgängig über die Ebenen</b>							
Gruppe 1:	62	65	71	65	71	53	58
Gruppe 2:	64	64	74	57	65	39	56
Gruppe 3:	23	25	25	22	25	22	22
<b>T4-N durchgängig über die Ebenen</b>							
Gruppe 1:	24+1	25		19	20	10	18
Gruppe 2:	12	12	/	11	15	4	8
Gruppe 3:	8	8		5	8	5	4

Reaktion unabhängig der Versuchsgruppen: Trainingsfragebogen direkt nach dem Training (T1), 1-2 Monate später (T2) und 6 Monate später (T4); +1 bedeutet fehlender T2-Bogen, aber vorhandener T4-Bogen; Lernen: unterteilt in Wissen (Vorhandensein des Wissenstests, nur zu T1 und T2 möglich) und Einstellungen (Vorhandensein der T1-T4-Fragebögen); Verhalten: vorhandene Fahrprofilaten (Handy und Video) in der jeweiligen Schicht, Resultate: Aufteilung in Schichtfragebögen (ob mindestens ein Schichtende vorhanden ist) und EKGs



### 3.4.4 Beschreibung der Stichprobe

Tabelle 10 und Abbildung 6 zeigen die demographischen und gesundheitlichen Merkmale der Stichprobe, wie sie im T1 Fragebogen in der Vortestung erhoben wurden.

Tabelle 10: Demographische Daten der Gesamtstichprobe zu Studienbeginn

Abhängige Variable	N	Möglicher Bereich	M	Min	Max	SD
Alter	182	/	33.1	20	65	9.4
Rettungsdienstenerfahrung (in Monaten)	178	/	128.9	4	456	102.8
Organisationszugehörigkeit (in Monaten)	179	/	98.0	5	456	95.2
Anzahl Einsatzfahrten im Monat	148	/	48.5	2	155	28.8
Anzahl Führerscheine	180	[1,7]	4.3	2	7	1.7
Anzahl Führerschein-klassen (B/C1+ A/C/E)	180	[1,4]	2.5	1	4	0.9
Jahre Führerscheinbesitz	179	/	14.6	1	44	8.9
Jährliche Fahrleistung privat (in km)	161	/	18279	20	52500	10073
Jährliche Fahrleistung beruflich (in km)	151	/	15048	800	45000	9906
Jährliche Fahrleistung gesamt (in km)	163	/	31840	2250	71000	15074
Anzahl verkehrsrelevante Trainings	180	/	1.2	0	7	1.3
Vertragliche Wochen-Arbeitszeit im Hauptamt (in Stunden)	160	/	41.6	10	48	6.1
Tatsächliche Wochen-Arbeitszeit im Hauptamt (in Stunden)	154	/	46.9	11	72	8.8
Vertragliche Wochen-Arbeitszeit im Nebenamt (in Stunden)	19	/	35.8	4	55	10.4
Tatsächliche Wochen-Arbeitszeit im Nebenamt (in Stunden)	20	/	16.5	2.5	48	12.0
Sportliche Aktivität	99	[1,5]	2.9	1	5	0.9
Größe	100	/	1.77	1.55	1.93	0.09
Gewicht	100	/	81.7	51	120	16.3
BMI	139	/	25.7	18.0	36.8	4.2

Abkürzungen: N=Anzahl der Probanden, M=Mittelwert, Min=Minimum, Max=Maximum, SD=Standardabweichung

Insgesamt 21.9% der Teilnehmer waren weiblich. In einer hier nicht näher dargestellten Befragung der teilgenommenen Organisationen (Organisationsscreening) wurden Frauenanteile bei den Rettungsdienstmitarbeitern zwischen 10.6% und 32.8% Prozent gefunden, so dass davon ausgegangen werden kann, dass die Stichprobe mit Blick auf die Geschlechterverteilung repräsentativ für die teilnehmenden Organisationen ist. Die meisten Teilnehmer (88.5%) sind hauptamtlich im Rettungsdienst beschäftigt, gefolgt von ehrenamtlichen (7.1%) und nebenamtlichen (4.4%) Mitarbeitern. Dies spiegelt nicht die tatsächliche Verteilung in den Organisationen wider, nach deren Angaben durchschnittlich 51.6% Personen hauptamtlich angestellt sind. Dieser Anteil unterscheidet sich jedoch stark nach den Bundesländern. In Sachsen sind mit 83.7% deutlich mehr Personen hauptamtlich beschäftigt als in Bayern mit 40.8%. Die Gründe hierfür liegen zum großen Teil an den Einschlusskriterien, die eine Mindestzahl an monatlichen Einsatzstunden erfordern, die gerade für neben- und ehrenamtliche Einsatzkräfte oft nicht regelmäßig erreicht werden. Die durchschnittliche tatsächliche Arbeitszeit der hauptamtlich Beschäftigten ist mit 46.9 Stunden etwas höher als der deutsche Durchschnitt von 43.5 Stunden, die der nicht hauptamtlichen ist mit 16.5 Stunden geringer als der Schnitt von 23.1 Stunden (BAuA, 2016). Das Alter und die Erfahrung im Rettungsdienst sowie im Straßenverkehr zeigen eine hohe Spannweite. Allerdings weisen die Teilnehmer im Vergleich zur

Allgemeinbevölkerung eine relativ hohe Fahrerfahrung auf. Alle Probanden besitzen mindestens einen C1-Führerschein, da dieser Voraussetzung für die meisten RTW ist. Laut Angaben des Kraftfahrt-Bundesamt liegt der Bestand an LKW-Führerscheinen (C1, C1E, C und CE) in Deutschland zum 1.1.2017 bei ca. 36.4% der gesamten vorhandenen Fahrerlaubnisklassen. Auch die durchschnittliche Fahrleistung von 31841 km liegt deutlich über dem deutschen Schnitt von 14015 km für PKW bzw. 18393 km für LKW (zwischen 3.5 und 6t). Die Anzahl von Personen mit bisherigen Verkehrsvergehen, bei denen es auch zu Punkten in Flensburg kam, scheint mit über der Hälfte der Probanden recht hoch zu sein. Bei vielen Probanden liegen die Verkehrsvergehen jedoch schon über zwei Jahre zurück. Die zum Erhebungszeitpunkt aktuellen Punkte liegen jedoch trotz erhöhter Fahrleistung mit 12.6% leicht unter dem Bevölkerungsdurchschnitt. Denn von den derzeit etwa 55 Millionen Führerscheinbesitzern gab es zum 1.1.2016 9.6 Millionen eingetragene Personen im Fahreignungsregister, was etwa 17.5% ausmacht. Diese Zahl ist zum 1.1.2017 aufgrund der verlängerten Tilgungsfrist auf 10.1 Millionen (ca. 18.3% der Führerscheinbesitzer) gestiegen.

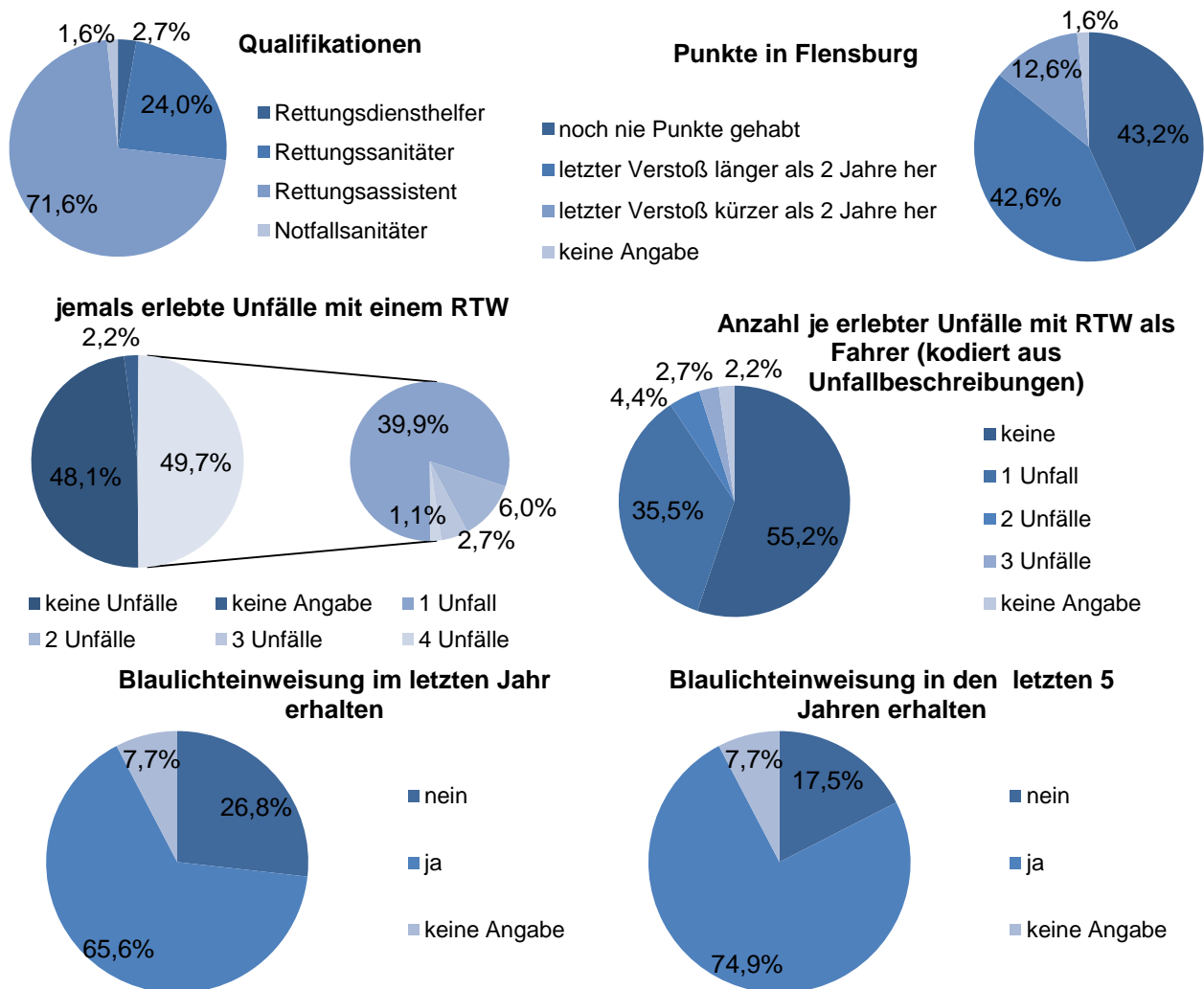


Abbildung 6: Auswahl kategorialer demographischer Daten (N=183 Probanden)

Die Anzahl erlebter Unfälle als Fahrer eines motorisierten Fahrzeugs in den letzten drei Jahren sowie die Anzahl von je erlebten Unfällen mit RTW als Fahrer (kategorisiert aus den Freitexten der Unfallbeschreibungen) erscheint in der Stichprobe auf den ersten Blick recht hoch. Fast die Hälfte der Probanden war innerhalb der letzten 3 Jahre als Fahrer eines motorisierten Fahrzeugs in einen Unfall verwickelt (47.5%). Die meisten erlebten Unfälle fanden dabei im dienstlichen Rahmen statt (51.7% nur dienstliche Unfälle, 24.1% eine Mischung aus dienstlichen und privaten Unfällen), obwohl der private Bereich die

etwas größere Fahrleistung aufweist. Im privaten Bereich haben 31 Probanden einen Unfall und 11 Probanden zwei Unfälle als Fahrer erlebt. Im dienstlichen Bereich wurden 45-mal ein Unfall, 17-mal zwei Unfälle, drei mal drei Unfälle und einmal 5 Unfälle selbst als Fahrer erlebt. Im Jahr 2016 fanden in Deutschland bei 725.8 Milliarden gefahrenen Kilometern (laut Angaben des Kraftfahrt-Bundesamts) und 2.6 Millionen Unfällen (Statistisches Bundesamt, 2017) etwa alle 280747 Kilometer ein Unfall statt. Im Schnitt alle 2.4 Millionen km fand ein Unfall mit Personenschaden statt. Mit einer durchschnittlichen Fahrleistung von 31841 km im Jahr wird man rein rechnerisch alle 8.8 Jahre in einen Unfall verwickelt. Damit ist die Anzahl der Probanden, die in den letzten Jahren Unfälle erlebt haben ebenso wie die Anzahl der Unfälle trotz der hohen Fahrleistung recht hoch.

Bei Unfällen im Rettungsdienst wurde insgesamt nach der Summe aller erlebten Unfälle gefragt, unabhängig davon, wie lange sie her sind oder ob die Probanden selbst Fahrer waren oder nicht. Knapp die Hälfte aller Probanden hat bereits einen Unfall mit einem RTW erlebt. Die Summe der Unfälle reichte dabei bis zu vier, die meisten erlebten jedoch bisher nur einen Unfall. Insgesamt 118 Unfälle wurden von den Probanden angegeben und beschrieben. Davon wurden 96 selbst als Fahrer erlebt. Da die Anzahl der als Fahrer erlebten Unfälle deutlich höher ist, als die als Beifahrer stellt sich bei der Gesamtzahl die Frage, ob die eigenen Unfälle besser erinnert werden und daher häufiger berichtet werden und andere vergessen und nicht angegeben wurden.

Glücklicherweise erlebte bisher keiner der Probanden einen Unfall mit Todesfolge in einem RTW. Insgesamt drei Unfälle wurden berichtet, bei denen zum Teil schwere Verletzungen der Beteiligten die Folge waren, fünf weitere mit Leichtverletzten. Die meisten Unfälle wiesen also „lediglich“ Sachschäden auf (42mal Sachschaden über 1500 Euro, 46mal Sachschaden unter 1500 Euro).

65.5% der Probanden haben eine Belehrung hinsichtlich der gesetzlichen Regelungen innerhalb des letzten Jahres erhalten. Das bedeutet jedoch auch, dass viele diese nicht erhalten haben (26.8%). Von denen ohne Belehrung haben viele auch die Jahre davor keine Belehrung zu diesem Thema erhalten. Insgesamt 17.5% der Probanden haben in den letzten fünf Jahren keine Blaulichtbelehrung bekommen. Blaulichtbelehrungen fallen in den Organisationen sehr unterschiedlich aus. Einige erläutern die beiden relevanten Paragraphen der Straßenverkehrsordnung andere besprechen Unfälle innerhalb der Organisation, deren Gründe und Fahrdaten.

Insgesamt ist die vorhandene Stichprobe hinsichtlich Alter und Geschlecht eine recht gute Abbildung der tatsächlichen Rettungsdienstmitarbeiter. Es wird eine höhere Anzahl hauptamtlicher Mitarbeiter betrachtet und die durchschnittliche tatsächliche Arbeitszeit übertrifft den Durchschnitt in Deutschland. Die Fahrerfahrung (Anzahl Führerscheine, durchschnittliche Jahreskilometer) und wahrscheinlich Unfälle liegen über, Verkehrsverstöße leicht unter dem Bevölkerungsdurchschnitt.

### **3.4.5 Verteilung der Kontrollvariablen in den Gruppen**

Das Studiendesign sieht, wie häufig bei Evaluationsstudien in der Praxis, aus organisatorischen Gründen keine Randomisierung der Teilnehmer auf die Studiengruppen vor. Dennoch wurde versucht, die Teilnehmer und Teilnehmerinnen möglichst gleichmäßig zu verteilen, um Unterschiede zwischen den Gruppen eindeutig auf die Evaluation und nicht auf etwaige Unterschiede in Drittvariablen zurückführen zu können. Tabelle 11 bis Tabelle 13 zeigen daher die Verteilung potenziell relevanter Merkmale in den Studiengruppen zu Studienbeginn für die Bereiche allgemeiner (berufsbezogener) demographischer, verkehrsrelevanter und gesundheitlicher Aspekte. Merkmale, die nicht gleichverteilt sind, sind in den weiteren Analysen als Kontrollvariablen zu berücksichtigen, während Merkmale, deren Ausprägungen in den drei Studiengruppen nicht unterschiedlich sind, für das Evaluationsergebnis keine relevante Rolle spielen sollten und damit nicht in die varianzanalytischen und/oder regressionsanalytischen Auswertungen einfließen müssen.

Tabelle 11: Unterschiede in den (berufsbezogenen) demographischen Merkmalen aller aufgenommenen Teilnehmer zwischen den Gruppen zu Studienbeginn

Merkmal	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenkon- trolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
Geschlecht	/ <sup>2</sup>	/	/	1.41	p=.478 n.s.
Männer	65	58	20		
Frauen	14	19	7		
Alter	/	/	/	2.32	p=.101 n.s.
	M = 34.3 SD = 10.17	M = 31.4 SD = 8.01	M = 34.6 SD = 10.00		
Qualifikation	/	/	/	3.95	p=.708 n.s.
Rettungsdiensthelfer	3	2	0		
Rettungssanitäter	22	18	4		
Rettungsassistent	53	56	22		
Notfallsanitäter	1	1	1		
Dauer Beschäftigung im Rettungsdienst (Monate)	b	a	/	3.30	p=.039*
	M = 144.3 SD = 106.48	M = 106.0 SD = 87.93	M = 148.9 SD = 120.80		
Hilfsorganisation	/	/	/	4.89	p=.899 n.s.
BRK	34	29	10		
JUH	6	4	2		
ASB	7	7	4		
MHD	11	8	4		
DRK	10	18	4		
Andere Anbieter	11	11	3		
Hilfsorganisation nach Rechtsform	/	/	/	0.18	p=.914 n.s.
privatwirtschaftlich	11	11	3		
freigemeinnützig	68	66	24		
Dauer Beschäftigung in Hilfsorganisation (Monate)	/	/	/	1.98	p=.141 n.s.
	M = 105.4 SD = 99.86	M = 82.7 SD = 79.70	M = 120.3 SD = 116.35		
Beschäftigungsart	/	/	/	6.03	p=.187 n.s.
hauptberuflich	65	72	25		
nebenberuflich	6	1	1		
ehrenamtlich	8	4	1		
Hauptamt: Vertragliche Arbeitszeit pro Woche	/	/	/	0.02	p=.979 n.s.
	M = 41.5 SD = 7.15	M = 41.7 SD = 5.00	M = 41.6 SD = 6.23		
Hauptamt: Tatsächliche Arbeitszeit pro Woche	/	/	/	0.38	p=.638 n.s.
	M = 47.4 SD = 10.03	M = 47.0 SD = 8.00	M = 45.5 SD = 7.34		
Neben-/Ehrenamt: Ver- tragliche Arbeitszeit pro Woche	/	/	/	0.74	p=.491 n.s.
	M = 37.1	M = 30.3	M = 39.0		

Merkmal	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenkon- trolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
	SD = 8.43	SD = 17.52	SD = 1.41		
Neben-/Ehrenamt: Tatsächliche Arbeitszeit pro Woche	/	/	/	0.36	p=.705 n.s.
	M = 14.9 SD = 10.84	M = 20.4 SD = 17.76	M = 17.0 SD = 9.90		
Bundesland	/	/	/	3.37	p=.191 n.s.
Bayern	57	47	15		
Sachsen	22	30	12		

<sup>1</sup> n.s. = nicht signifikant, keine Unterschiede in der Verteilung des Merkmals über die drei Gruppen vorhanden, Signifikanzniveaus: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

<sup>2</sup> / = nicht signifikant im Einzelgruppenvergleich; Buchstaben geben die Gruppe an, zu der ein Unterschied in paarweisen Vergleichen auf einem 5%-Signifikanzniveau bestehen

Für die allgemeinen (berufsbezogenen) Variablen in Tabelle 11 zeigt sich, dass die Verteilung der Teilnehmer auf die Studiengruppen insgesamt sehr gleichmäßig gelungen ist. Lediglich für die Länge der Beschäftigung im Rettungsdienst findet sich ein signifikanter Effekt zwischen den Gruppen, der wesentlich dadurch bestimmt ist, dass die Wartegruppe (Gruppe 2) erst kürzer im Rettungsdienst arbeitet als die Interventionsgruppe (Gruppe 1). Dieser Unterschied scheint wesentlich darin begründet zu sein, dass die Wartegruppe durchschnittlich etwa drei Jahre jünger ist als die Interventionsgruppe – auch wenn dieser Unterschied konventionelle Signifikanzniveaus knapp nicht erreicht. Beide Variablen korrelieren mit  $r = .828$  ( $p = .000$ ) und haben damit nahezu 70% gemeinsame Varianz. Dies kann bei gemeinsamer Aufnahme in Analysen zur Varianzinflationierung und damit zu nicht validen Ergebnissen führen. Vor diesem Hintergrund wird – als die „üblichere“ Kontrollvariable - Alter in die allgemeinen Analysen aufgenommen und Beschäftigungszeit im Rettungsdienst nur für spezifische Auswertungen herangezogen.

Tabelle 12: Unterschiede in den verkehrsrelevanten Merkmalen aller aufgenommenen Teilnehmer zwischen den Gruppen zu Studienbeginn

Merkmal	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenkon- trolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
Anzahl Führerscheine	/ <sup>2</sup>	/	/	1.15	p=.319 n.s.
	M = 4.18 SD = 1.60	M = 4.30 SD = 1.72	M = 4.74 SD = 1.58		
Führerscheinklassen	/	/	/	6.26	p=.401 n.s.
Nur B und C1	14	11	3		
B und C1 + 1 (A C E)	27	30	6		
B und C1 + 2 (A C E)	27	21	13		
B und C1 + 3 (A C E)	9	14	5		
Jahre Führerscheinbesitz	b	a	/	3.45	p=.034*
	M = 16.1 SD = 9.81	M = 12.5 SD = 7.28	M = 15.9 SD = 9.77		
Erlaubnis zur Fahrgastbeförderung	/	/	/	5.53	p=.068 <sup>+</sup>
ja	12	22	3		
nein	62	53	23		

Abschlussbericht „Achtung Blaulicht“

Merkmal	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenkon- trolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
Jährliche Fahrleistung privat in km (alle Fahrzeugklassen)	/	/	/	0.45	p=.637 n.s.
	M = 19176 SD = 10256.6	M = 17627 SD = 9703.4	M = 17674 SD = 10755.6		
Jährliche Fahrleistung beruf- lich in km (alle Fahrzeugklassen)	/	/	/	0.35	p=.708 n.s.
	M = 15422 SD = 10415.9	M = 15274 SD = 9478.7	M = 13542 SD = 9940.0		
Jährliche Fahrleistung gesamt in km (bereinigt, alle Fahrzeugklassen)	/	/	/	0.38	p=.683 n.s.
	M = 33071 SD = 15380.4	M = 31088 SD = 14894.9	M = 30695 SD = 15117.6		
Punkte Fahreignungsregister (jemals)	/	/	/	1.13	p=.583 n.s.
ja	46	42	13		
nein	31	34	14		
Wenn Punkte, dann länger als 2 Jahre zurück?	/	/	/	0.62	p=.764 n.s.
ja	34	34	10		
nein	12	8	3		
Unfall (auch leichte) inner- halb der letzten 3 Jahre?	/	/	/	0.93	p=.650 n.s.
ja	37	37	16		
nein	37	39	11		
Anzahl Unfälle innerhalb der letzten drei Jahre?	/	/	/	4.10	p=.393 n.s.
Keine Unfälle	37	40	11		
Summe dienstliche Unfälle	46	32	15		
Summe private Unfälle	19	26	8		
Jemals in Unfall mit Ret- tungswagen involviert?	/	/	/	0.93	p=.652 n.s.
ja	38	37	16		
nein	38	39	11		
Anzahl der Unfälle mit RTW als Fahrer	/	/	/	0.79	p=.454 n.s.
	M = 0.61 SD = 0.80	M = 0.46 SD = 0.62	M = 0.56 SD = 0.70		
Schwere der Unfälle mit RTW als Fahrer (Rating) <sup>3</sup>	/	/	/	3.36	p=.762 n.s.
Sachschäden unter 1.500 €	19	16	8		
Sachschäden über 1.500 €	20	17	5		
leichte Verletzungen	2	1	2		
schwere Verletzungen	1	1	0		
Stunden pro Monat als Fahrer auf RTW	/	/	/	1.27	p=.284 n.s.
	M = 77.1 SD = 45.58	M = 87.7 SD = 53.23	M = 72.3 SD = 49.0		

Abschlussbericht „Achtung Blaulicht“

Merkmal	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenkon- trolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
Stunden pro Monat als Beifahrer auf RTW	/	/	/	0.50	p=.608 n.s.
	M = 70.6 SD = 43.39	M = 64.4 SD = 46.08	M = 61.1 SD = 45.01		
Anzahl der Einsatzfahrten mit Sondersignal als Fahrer pro Monat	/	/	/	0.32	p=.724 n.s.
	M = 48.1 SD = 30.69	M = 52.4 SD = 28.11	M = 48.3 SD = 29.54		
„Freude am Fahren mit Sondersignal“ (Fahrer)	/	/	/	1.39	p=.253 n.s.
	M = 3.49 SD = 0.94	M = 3.76 SD = 0.96	M = 3.69 SD = 1.08		
Beanspruchung durch Fah- ren mit SoSi (Fahrer)	/	/	/	1.79	p=.170 n.s.
	M = 3.07 SD = 1.17	M = 2.78 SD = 1.00	M = 2.67 SD = 1.29		
„Freude am Fahren mit Sondersignal“ (Beifahrer)	/	/	/	0.58	p=.563 n.s.
	M = 2.88 SD = 0.95	M = 3.02 SD = 0.90	M = 2.79 SD = 1.06		
Beanspruchung durch Fah- ren mit SoSi (Beifahrer)	/	/	/	0.97	p=.380 n.s.
	M = 3.00 SD = 1.07	M = 2.75 SD = 1.06	M = 2.71 SD = 1.27		
Anzahl verkehrsrelevanter Fortbildungen	/	/	/	5.92	p=.436 n.s.
Keine	28	22	9		
1 - 2	35	46	14		
3 - 4	11	7	2		
5 und mehr	3	1	2		
Blaulichteinweisung inner- halb der letzten fünf Jahre?	/	/	/	0.03	p=.984 n.s.
nein	13	14	5		
ja	58	58	21		

<sup>1</sup> n.s. = nicht signifikant, keine Unterschiede in der Verteilung des Merkmals über die drei Gruppen vorhanden, Signifikanzniveaus: <sup>†</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

<sup>2</sup> / = nicht signifikant im Einzelgruppenvergleich; Buchstaben geben die Gruppe an, zu der ein Unterschied in paarweisen Vergleichen auf einem 5%-Signifikanzniveau bestehen

<sup>3</sup> Summe der Ratings aus qualitativen Beschreibungen der Unfälle; nur beschriebene Schweregrade dargestellt, z.B. kein Unfall mit Todesfolge genannt

Ähnliches zeigt sich im Bereich der verkehrsrelevanten Merkmale (vgl. Tabelle 12) für die Jahre des Führerscheinbesitzes. Erneut haben entsprechend dem Altersunterschied die Teilnehmer der Wartegruppe (Gruppe 2) signifikant und durchschnittlich etwa dreieinhalb Jahre kürzer ihren Führerschein als die Teilnehmer der Interventionsgruppe (Gruppe 1). Hier beträgt die Korrelation zwischen beiden Variablen sogar  $r = .968$  ( $p = .000$ ) mit knapp 94% gemeinsamer Varianz, so dass eine gemeinsame Aufnahme in eine Varianz- oder Regressionsanalytische Analyse ausgeschlossen ist. Daher wird mit den Jahren des Führerscheinbesitzes genauso verfahren wie mit der Beschäftigungszeit im Rettungsdienst. Weitere systematische

Unterschiede zwischen den Studiengruppen im Bereich verkehrsrelevanter Merkmale finden sich nicht; allerdings besteht für das Vorhandensein einer Erlaubnis zur Fahrgastbeförderung ein tendenzieller Unterschied bei einer erhöhten Irrtumswahrscheinlichkeit von 10%. Da zudem für diese Variable keiner der einfachen Gruppenvergleiche signifikant wird, wird von einer Kontrolle in den Hauptanalysen abgesehen.

Da auf der Evaluationsebene der Resultate auch gesundheitliche Parameter herangezogen werden, werden in Tabelle 13 die Ausprägungen einiger potenziell relevanter „medizinischer“ Merkmale in den Studiengruppen dargestellt. Zu beachten ist, dass nicht alle Personen diese Daten zu Studienbeginn angeben konnten, so dass die Stichproben hier insgesamt kleiner sind. Es zeigen sich keine systematischen Unterschiede zwischen den Gruppen, allerdings ist auffällig, dass vor allem für die diagnostizierten Herz-Kreislauf-Erkrankungen und regelmäßige Medikationen vergleichsweise hohe Anteile der Teilnehmer betroffen sind. Ebenso liegt der mittlere BMI in allen drei Gruppen über 25, der Wert nach dem gemäß Weltgesundheitsorganisation (WHO Global Data Base on Body-Mass-Index; <http://apps.who.int/bmi/index.jsp>) von Übergewicht bzw. Präadipositas gesprochen wird. 51% der zu Studienbeginn auf die Frage antwortenden Teilnehmer gaben an, regelmäßig oder gelegentlich zu rauchen.

Tabelle 13: Unterschiede in den medizinischen Merkmalen aller aufgenommenen Teilnehmer zwischen den Gruppen zu Studienbeginn

Merkmalsname	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenkon- trolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
Raucherstatus	/ <sup>2</sup>	/	/	5.11	p=.546 n.s.
noch nie	11	19	4		
nicht mehr	4	7	2		
gelegentlich	7	4	2		
Ja	19	13	4		
Herz-Kreislauf-Erkrankung diagnostiziert?	/	/	/	1.30	p=.537 n.s.
ja	15	11	3		
nein	61	63	24		
Regelmäßige Medikation (ohne Kontrazeptiva)	/	/	/	1.83	p=.408 n.s.
ja	11	9	6		
nein	68	68	21		
Körperliche Aktivitäten (für mindestens 20 Minuten) körperliche Aktivitäten?	/	/	/	10.03	p=.260 n.s.
täglich	2	2	0		
3-6x pro Woche	12	13	9		
1-2x pro Woche	17	15	4		
seltener	10	11	0		
nie	2	2	0		
Größe in m	/	/	/	0.32	p=.724 n.s.
	M = 1.77 SD = 0.09	M = 1.78 SD = 0.07	M = 1.78 SD = 0.05		
Gewicht in kg	/	/	/	0.20	p=.818 n.s.
	M = 82.0 SD = 15.65	M = 80.7 SD = 16.33	M = 79.6 SD = 13.77		



Merkmal	Gruppe 1 (a): Interventions- gruppe, n ≤ 79	Gruppe 2 (b): Wartegruppe, n ≤ 77	Gruppe 3 (c): Methodenk- ontrolle, n ≤ 27	χ <sup>2</sup> oder F- Statistik	Unter- schied <sup>1</sup>
BMI	/	/	/	0.75	p=.474 n.s.
	M = 26.2 SD = 4.37	M = 25.4 SD = 4.12	M = 25.1 SD = 3.94		

<sup>1</sup> n.s. = nicht signifikant, keine Unterschiede in der Verteilung des Merkmals über die drei Gruppen vorhanden, Signifikanzniveaus: †p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

<sup>2</sup> / = nicht signifikant im Einzelgruppenvergleich; Buchstaben geben die Gruppe an, zu der ein Unterschied in paarweisen Vergleichen auf einem 5%-Signifikanzniveau bestehen

Im Vergleich zu Bevölkerungsdaten, nach denen gemäß der GEDA-Studie 2009 (Robert-Koch-Institut, 2010) 30% der erwachsenen Bevölkerung raucht und dieser Trend weiterhin rückläufig ist, ist dieser Wert ebenfalls deutlich erhöht. Zusammengenommen erscheint die Stichprobe damit insgesamt eher schlechter ausgeprägte gesundheitsbezogene Werte zu haben.

### 3.5 Beschreibung der Ausgangswerte vor dem Training

Ein häufig nicht näher betrachtetes Problem bei der Evaluation von Maßnahmen oder Interventionen ist die Frage, ob das Ausgangsniveau der Teilnehmer in den interessierenden Variablen nicht bereits so hoch ist, dass eine weitere Verbesserung durch ein Training kaum oder gar nicht mehr möglich ist („Deckeneffekte“). Um dieses Phänomen in der vorliegenden Untersuchung zu berücksichtigen, werden für die folgende Analyse stichprobenunabhängige Werte für gute Ausprägungen in den Zielmerkmalen der Ebenen Lernen (Wissen und Einstellungen) sowie Verhalten (Fahrprofile) definiert. Die Reaktionsebene kann zum ersten Messzeitpunkt nicht eingeschlossen werden, da weder das Training zu dieser Zeit bereits stattgefunden hat noch eine wünschenswerte Ausprägung der Reaktion definiert werden kann. Die Resultateebene wird ausgeschlossen, da Unterschiede vor allem in den Gesundheitsaspekten aufgrund der fehlenden Vorgeschichte der Teilnehmer zu T1 nicht mit geringerer Beanspruchung bei Fahrten in Verbindung gebracht werden können.

Für das Ausgangs-Wissensniveau wird ein Wissenstestwert größer 12 als sehr guter Kenntnisstand angenommen. Der Wissenstest mit einem Maximalwert von 24 Punkten erfragt neben generellen Kenntnissen auch spezifische Wissensinhalte aus dem Training, so dass eine richtige Beantwortung von mehr als 50 Prozent bereits als positiv betrachtet wird. Im Bereich der Einstellungen und Risikowahrnehmung wird als theoretischer Grenzwert in Abhängigkeit vom Skalenniveau entweder das oberste/unterste Drittel oder das oberste/unterste Viertel der Skala betrachtet. Für die Einstellungen sind das die folgenden, jeweils vierstufig erfassten, wünschenswerten verkehrsbezogenen „Persönlichkeits-eigenschaften“: Extraversion < 2, Sensation Seeking < 2, Gewissenhaftigkeit > 3, Unverträglichkeit < 2 und Reaktanz < 2. Von einem sehr guten Ausgangsniveau bzw. einer Deckenausprägung in diesen Einstellungen, die kaum noch verbessert werden kann, wird ausgegangen, wenn alle fünf Aspekte den angegebenen Regeln entsprechen. Für die fünfstufig erfasste Risikowahrnehmung werden die folgenden Ausprägungen als wünschenswert und nur wenig verbesserbar angesehen: Risikowahrnehmung bei schweren Verstößen > 4, Risikowahrnehmung bei mittelschweren Verstößen > 4, Risikowahrnehmung bei leichten Verstößen > 4 sowie einem Risk seeking < 2. Auch hier wird dann von einem sehr guten, kaum noch steigerungsfähigen Ausgangsniveau ausgegangen, wenn alle fünf Bedingungen gegeben sind.

Um ein sehr gutes, nur wenig verbesserungsfähiges Fahrverhalten zu definieren, werden theoretisch abgeleitete Trigger für starke Beschleunigungen, Vollbremsungen und zu starkes Rechts- oder Links-einlenken („kritische Manöver“) verwendet sowie eher defensive Durchschnittsgeschwindigkeiten bei Hinfahrten mit Sondersignal herangezogen. Für Rückfahrten mit Sondersignal werden identische Trigger

wie bei Hinfahrten verwendet. Um individuellen Fahr-Gewohnheiten zu entsprechen, wird für die Rückfahrten eine Geschwindigkeit als positiv angenommen, die maximal gleich schnell ( $\Delta$ Geschwindigkeit  $\leq 1$  km/h) oder langsamer ist als die Hinfahrt. Zusammengenommen sollen diese Kriterien eine Darstellung für gleichmäßiges, ruhiges und eher defensives Fahrverhalten sein. Wo immer möglich, wird bei den Berechnungen der Fahrzeugtyp (RTW vs. NEF) berücksichtigt.

Als Trigger wurden definiert: eine positive Längsbeschleunigung von  $5,89 \text{ m/s}^2$ , eine negative Längsbeschleunigung (Vollbremsung) von  $-3 \text{ m/s}^2$  für RTW (basierend auf RTW-Messungen, vgl. Abschnitt 0) und von  $-5,89 \text{ m/s}^2$  für NEF sowie eine Querschleunigung von  $\pm 6,87 \text{ m/s}^2$ . Für letzteren Trigger liegen leider keine gesonderten Werte für RTW und NEF vor. Da jedoch die verwendeten NEF- und RTW-Fahrzeugtypen zwischen den untersuchten Bundesländern Sachsen und Bayern im Hinblick auf potenzielle „Kipprisiken“ deutlich variieren, wäre hier eine genauere Aufteilung vermutlich auch nicht zielführend.

Zur Ermittlung angemessener, stichprobenunabhängiger Durchschnittsgeschwindigkeiten wurde zunächst von den zulässigen Höchst-/Richtgeschwindigkeiten plus einem „Sondersignal“-Aufschlag von 5% ausgegangen. Zudem wurde, wie in Gauger (2011) für Baden-Württemberg berichtet, eine durchschnittliche Einsatzfahrt mit 80% Innerorts-Fahrten, 10% Autobahnfahrten und 10% Land- und Bundesstraßen angenommen. Mit diesen Setzungen ergeben sich wünschenswerte durchschnittliche Geschwindigkeiten für RTW (mit Gesamtgewichten zwischen 3,5 und 7,5 t) von:  $((50 \text{ km/h}+5\%)*80)+((80 \text{ km/h}+5\%)*10)+((80 \text{ km/h}+5\%)*10)/100 = 58,8 \text{ km/h}$  und für NEF (als PKW) von:  $((50 \text{ km/h}+5\%)*80)+((100 \text{ km/h}+5\%)*10)+((130 \text{ km/h}+5\%)*10)/100 = 66,15 \text{ km/h}$ . Diese Werte sind in etwa vergleichbar mit den von Gauger (2011) und Steinvoord (2012) genannten möglichen Durchschnittsgeschwindigkeiten bei Nutzung des von Schmiedel, Betzler und Behrendt (2004) beschriebenen Verfahrens zur Ermittlung von Planungsgeschwindigkeiten.

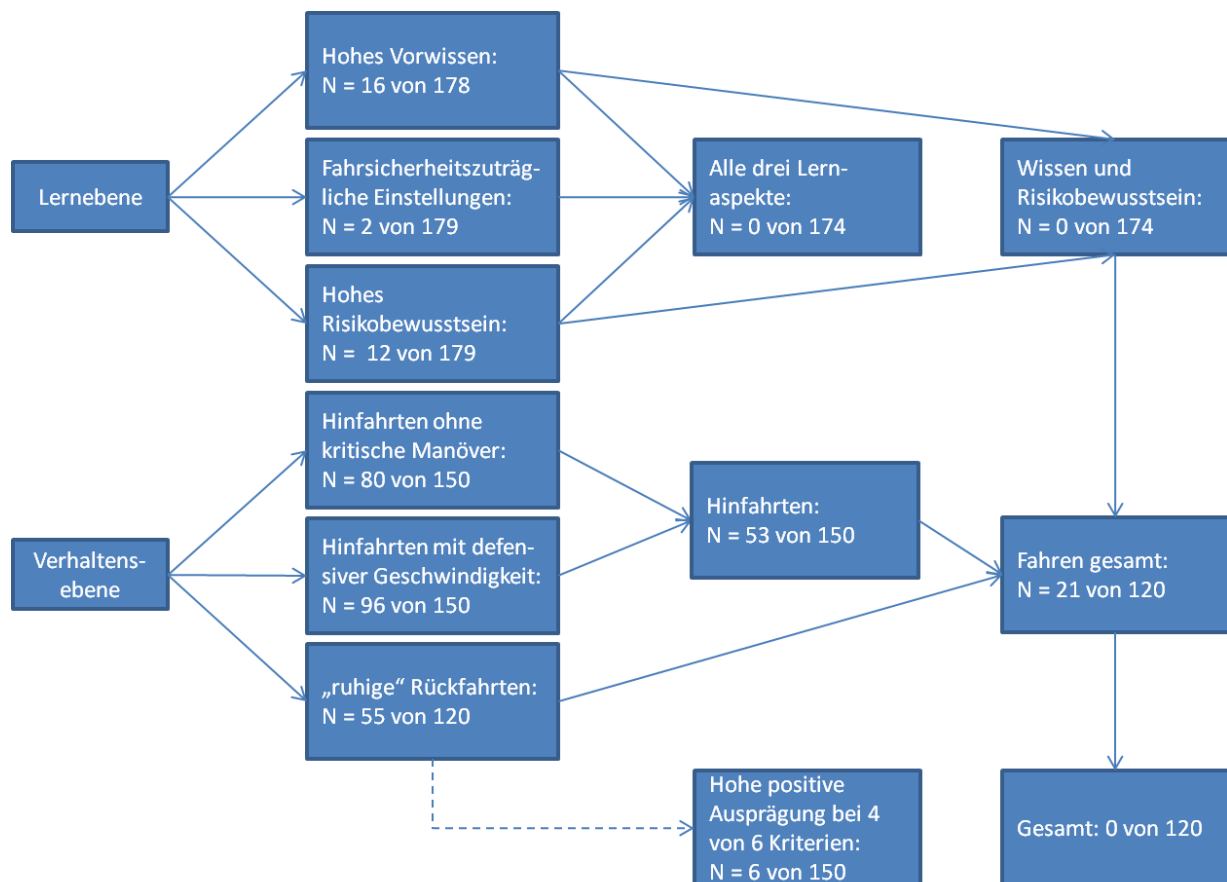


Abbildung 7: Personen mit potenziellen Deckeneffekten zum ersten Messzeitpunkt

Abbildung 7 zeigt die jeweilige Anzahl von Personen, die diese Bedingungen zu Beginn der Studie erfüllen und damit über eine (sehr) gute Ausprägung der im Verlauf der Studie zu trainierenden Inhalte verfügen. Insgesamt fällt auf, dass insgesamt nur wenige Teilnehmer hohe Ausprägungen bei den verkehrsrelevanten Wissens- und Einstellungsaspekten haben und keiner der Teilnehmer alle drei Kriterien erfüllt, während dies im Bereich des Fahrverhaltens insgesamt etwas günstiger aussieht und hier 21 Personen die Kriterien für gleichmäßiges, ruhiges und eher defensives Fahrverhalten erfüllen. Für einen potenziellen Deckeneffekt im Hinblick auf das gesamte Training kann – mit vier von sechs erfüllten Kriterien – nur von sechs Personen ausgegangen werden. Bei diesen sechs Teilnehmern (zwei aus der Interventionsgruppe, vier aus der Wartegruppe) handelt es sich um fünf Männer und eine Frau mit einem durchschnittlichen Alter von 31.5 Jahren, vier stammen aus Sachsen, zwei aus Bayern, alle sechs kommen aus unterschiedlichen Rettungsdienstorganisation. Zwei dieser bereits zu Beginn der Studie erfolgreichen Teilnehmer sind Rettungssanitäter, vier sind Rettungsassistenten, alle arbeiten hauptamtlich in dieser Tätigkeit. Leichte Unfälle mit RTW haben drei der sechs erlebt, zwei davon als Fahrer. Vier haben noch nie Punkte im Fahreignungsregister gehabt, bei den zwei anderen sind die Punkte länger als zwei Jahre her. Obwohl sich aufgrund der extremen Größenunterschiede in den Teilstichproben eine statistische Analyse verbindet, scheinen diese Studienteilnehmer im Hinblick auf Ausbildung und RTW-Unfälle vergleichbar mit der Gesamtstichprobe, sind tendenziell zwei Jahre jünger und haben mit 17% einen etwas geringeren Frauenanteil als die Gesamtstichprobe (22%). Hauptamtlich Tätige sind mit 100% überrepräsentiert (Gesamtstichprobe etwa 89%). Besonders auffällige Unterschiede finden sich im Hinblick auf Bundesland und Fahreignungsregister: Während in der Stichprobe insgesamt nur etwa 35% sächsische Teilnehmer sind, finden sich unter den sechs bereits guten Fahrern Zweidrittel Sachsen. Ebenso haben Zweidrittel der sehr guten sechs Personen noch nie Punkte im Fahreignungsregister gehabt, während dies in der Gesamtstichprobe nur auf etwa 44% zutrifft.

Tabelle 14 beschreibt ausgewählte Merkmale der 21 Teilnehmer, die auf der Verhaltensebene bereits am Beginn der Studie gut abgeschnitten haben, im Vergleich zu den anderen Personen, von denen Fahrprofilaten zum ersten Messzeitpunkt vorliegen. Bei keiner der demographischen Variablen finden sich systematische Unterschiede zwischen diesen Gruppen. Tendenziell (mit knapp über 10% Irrtumswahrscheinlichkeit) finden sich, wie bei den Personen mit vier von sechs erfüllten Kriterien, mehr Teilnehmer aus Sachsen unter den „guten Fahrern“ und weniger Personen mit Punkten im Fahreignungsregister.

Tabelle 14: Teilnehmer mit guten Fahrprofilen zu Studienbeginn im Vergleich zu anderen Teilnehmern: Demographische Aspekte

Merkmal	„Gute Fahrer“ zu T1, n = 21	„Andere Fahrer“ zu T1 n ≤ 129	χ2 oder T-Statistik	Unterschied <sup>1</sup>
Geschlecht			0.01	p=.586 n.s.
Männer	17	103		
Frauen	4	26		
Alter	M = 30.9, SD = 7.70	M = 33.1, SD = 9.33	-1.95	p=.294 n.s.
Qualifikation			3.11	p=.384 n.s.
Rettungsdiensthelfer	0	5		
Rettungssanitäter	6	32		
Rettungsassistent	14	91		
Notfallsanitäter	1	1		
Dauer Beschäftigung im Rettungsdienst (Monate)	M = 117.4 SD = 86.65	M = 126.6 SD = 100.09	-0.39	p=.695 n.s.
Hilfsorganisation nach Rechtsform	/	/	0.20	p=.652 n.s.
privatwirtschaftlich	5	17		
freigemeinnützig	16	112		

Abschlussbericht „Achtung Blaulicht“

Merkmal	„Gute Fahrer“ zu T1, n = 21	„Andere Fahrer“ zu T1 n ≤ 129	χ <sup>2</sup> oder T-Statistik	Unterschied <sup>1</sup>
Beschäftigungsart	/	/	3.04	p=.223 n.s.
hauptberuflich	19	113		
nebenberuflich	2	5		
ehrenamtlich	0	11		
Bundesland	/	/	2.24	p=.107 n.s.
Bayern	11	89		
Sachsen	10	40		
Anzahl Führerscheinklassen	M = 3.95, SD = 1.28	M = 4.31, SD = 1.72	-0.91	p=.364 n.s.
Jahre Führerscheinbesitz	M = 12.8; SD = 7.12	M = 14.6; SD = 8.92	-0.87	p=.384 n.s.
Erlaubnis zur Fahrgastbeförderung	/	/	0.54	p=.315 n.s.
ja	6	26		
nein	15	96		
Jährliche Fahrleistung gesamt in km (alle Fahrzeugklassen)	M = 29565 SD = 15129.7	M = 32591 SD = 15194.3	-0.80	p=.423 n.s.
Punkte Fahreignungsregister	/	/	2.25	p=.104 n.s.
ja	9	76		
nein	12	50		
Jemals in Unfall mit Rettungswagen involviert?	/	/	0.06	p=.500 n.s.
ja	10	63		
nein	11	62		
Anzahl der Unfälle mit RTW als Fahrer	M = 0.52, SD = 0.75	M = 0.54, SD = 0.72	-0.12	p=.907 n.s.
Stunden pro Monat als Fahrer auf RTW	M = 95.6 SD = 51.81	M = 80.2 SD = 49.49	1.28	p=.204 n.s.
„Freude am Fahren mit Sondersignal“ (Fahrer)	M = 3.81 SD = 0.87	M = 3.60 SD = 0.96	0.94	p=.350 n.s.
Beanspruchung durch Fahren mit SoSi (Fahrer)	M = 2.76 SD = 1.0	M = 2.93 SD = 1.10	-0.67	p=.505 n.s.
Anzahl verkehrsrelevanter Fortbildungen	/	/	0.82	p=.843 n.s.
Keine	8	41		
1 - 2	12	64		
3 - 4	1	7		
5 und mehr	0	4		
Blaulichteinweisung innerhalb der letzten fünf Jahre?	/	/	0.02	p=.554 n.s.
nein	4	22		
ja	16	95		

<sup>1</sup> n.s. = nicht signifikant, Signifikanzniveaus: \*p ≤ .10; \*\*p ≤ .05; \*\*\*p ≤ .01; \*\*\*\*p ≤ .001

Um zu analysieren, ob möglicherweise unterschiedliche Arbeitsbedingungen zu besserem oder nicht so gutem Fahrverhalten führen, werden in Tabelle 15 analog zu den demographischen Daten, Gruppenunterschiede zwischen den guten und den anderen Fahrern zum ersten Messzeitpunkt präsentiert. Erneut finden sich zwischen den guten und anderen Fahrern keinerlei Unterschiede in den Arbeitsbedingungen mit der (kontraintuitiven) Ausnahme, dass gute Fahrer mehr Arbeitsunterbrechungen berichten als die anderen Fahrer. Vor dem Hintergrund der nur geringen Anzahl von Teilnehmern mit bereits insgesamt oder

fahrbezogenen guten Werten zu T1 und ihrer gleichmäßigen Verteilung auf Interventions- und Wartegruppe, wird von einer gesonderten Kontrolle dieses Aspekts in den folgenden Analysen abgesehen.

Tabelle 15: Teilnehmer mit guten Fahrprofilen zu Studienbeginn im Vergleich zu anderen Teilnehmern: Arbeitsbedingungen

Merkmal	„Gute Fahrer“ zu T1, n = 21		„Andere Fahrer“ zu T1 n ≤ 129		T-Statistik <sup>1</sup>	Unterschied <sup>2</sup>
	M	SD	M	SD		
<b>Anforderungen</b>						
Geistige Anforderungen	4.27	0.558	4.14	0.571	0.99	.323 n.s.
Lernerfordernisse	4.29	0.530	4.08	0.644	1.37	.174 n.s.
Qualifizierungsmöglichkeiten	3.98	0.826	3.81	0.798	0.95	.346 n.s.
Emotionale Anforderungen	2.85	0.756	2.81	0.676	0.22	.824 n.s.
<b>Ressourcen</b>						
Zeitliche Transparenz/ Vorhersehbarkeit	3.01	0.763	3.02	0.658	-0.04	.969 n.s.
Tätigkeitsspielräume	3.26	0.683	3.21	0.628	0.33	.744 n.s.
Arbeitszeitautonomie	2.25	0.650	2.31	0.833	-0.31	.757 n.s.
Vorgesetztenfeedback	2.45	1.071	2.29	0.996	0.68	.501 n.s.
Partizipationschancen	2.30	0.918	2.27	0.741	0.18	.859 n.s.
Räumliche Ressourcen	2.87	0.895	3.22	1.175	-1.57	.126 n.s.
Materielle Ressourcen	3.49	0.785	3.75	0.749	-1.41	.162 n.s.
Rollenklarheit	4.15	0.812	4.13	0.546	0.14	.890 n.s.
Führungsqualität	2.89	1.113	2.97	0.913	-0.36	.717 n.s.
Interne Kooperation	3.80	0.835	3.87	0.663	-0.43	.666 n.s.
Externe Kooperation	3.79	0.470	3.73	0.486	0.50	.618 n.s.
Soziale Unterstützung Vorgesetzte	3.01	1.125	3.23	1.025	-0.87	.387 n.s.
Soziale Unterstützung Kollegen	3.71	0.830	3.68	0.879	0.15	.883 n.s.
<b>Stressoren</b>						
Organisationale Stressoren	3.18	0.765	3.05	0.740	0.74	.461 n.s.
Patientenbezogene Stressoren	3.09	0.793	2.87	0.671	1.32	.187 n.s.
Zeitdruck bei unspezifischen Festlegungen	2.98	0.942	2.66	0.756	1.45	.159 n.s.
Zeitdruck bei spezifischen Festlegungen	2.27	0.729	2.27	0.538	-0.03	.976 n.s.
Widersprüchliche Aufträge	2.30	0.753	2.34	0.687	-0.30	.763 n.s.
Informativische Erschwernisse	3.14	0.730	3.14	0.643	-0.03	.974 n.s.
Arbeitsunterbrechungen	2.71	0.634	2.46	0.525	2.02	.045*
Qualitätseinbußen	2.25	0.852	2.21	0.697	0.21	.831 n.s.
Ungünstige Arbeitsumgebung (Einsatzort)	3.62	0.724	3.51	0.573	0.66	.514 n.s.
Ungünstige Arbeitsumgebung (RTW)	2.60	0.676	2.53	0.554	0.53	.596 n.s.
Strafender Umgang mit Fehlern	2.37	0.955	2.50	0.774	-0.69	.491 n.s.
Arbeitsplatzunsicherheit	1.68	0.671	1.73	0.615	-0.32	.748 n.s.

<sup>1</sup> T-Test für unverbundene Stichproben (korrigiert für Varianzungleichheit);

<sup>2</sup> n.s. = nicht signifikant, Signifikanzniveaus: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

## 4. Ergebnisse des Gesamtvorhabens

In diesem Abschnitt werden alle Ergebnisse zur Trainingsevaluation vorgestellt. Zunächst werden die durchgeführten Trainings in ihrer Zusammensetzung vorgestellt. Im Anschluss daran werden die Haupteffekte in allen Ebenen analysiert. Dafür werden die Gruppen zu den Zeitpunkten T1 und T2 verglichen (Hauptanalyse). Im nächsten Schritt erfolgt die Überprüfung, ob die reine Messung sich auf die abhängigen Variablen ausgewirkt hat. Zudem wird empirisch getestet, ob die t2 und t3 Messwerte der Gruppe 2, zwischen denen diese Wartegruppe das Training erhielt, in t1' und t2' überführt werden können, so dass für alle drei Gruppen hypothetisch angenommen werden kann, dass das Training zwischen den gleichen Messzeitpunkten stattfand, und damit Veränderungen über die Messungen t1, t2 und t4 mit einer einzigen Gruppe und damit höherer statistischer Power berechnet werden können.

Um weitere Effekte des Trainings überprüfen zu können, werden zusätzlich zu den Haupteffekten die Kontrollgruppe zwischen den Zeitpunkten T1 und T3 (Replizierbarkeit) und die Follow-up-Messungen (Nachhaltigkeit) analysiert. Um Hinweise darauf zu erhalten, ob mögliche Effekte nur durch Drop-Outs auftreten, werden abschließend Effekte auch mit allen vorliegenden Daten, d.h. mittels ungepaarter Stichproben, überprüft.

### 4.1 Beschreibung der Trainingszusammensetzungen

Da aufgrund organisatorischer Voraussetzungen (u.a. notwendige Freistellung der Mitarbeiter für die Trainings) eine zufällige oder stratifizierte Zuordnung der Studienteilnehmer zu den Trainings nicht möglich war und zudem nicht alle Trainingsgruppen mit den geplanten sechs Personen durchgeführt werden konnten, wird im Folgenden kurz beschrieben, wie die einzelnen Trainings zusammengesetzt waren. Insgesamt wurden 35 Trainings durchgeführt mit 166 Probanden und 10 weiteren Personen. Diese weiteren Personen wurden eingesetzt, um die Trainingsgrößen möglichst konstant zu halten und waren im wesentlichen Doktoranden und Praktikanten des Instituts, die alle auch Rettungsdiensterfahrung hatten. Die folgende Abbildung gibt einen Überblick über die Anzahl der Trainingsteilnehmer und die Zusammensetzung der Trainings hinsichtlich der Geschlechter. Entsprechend der geringeren Zahl an weiblichen Teilnehmern waren nahezu alle Trainingsgruppen männlich dominiert.

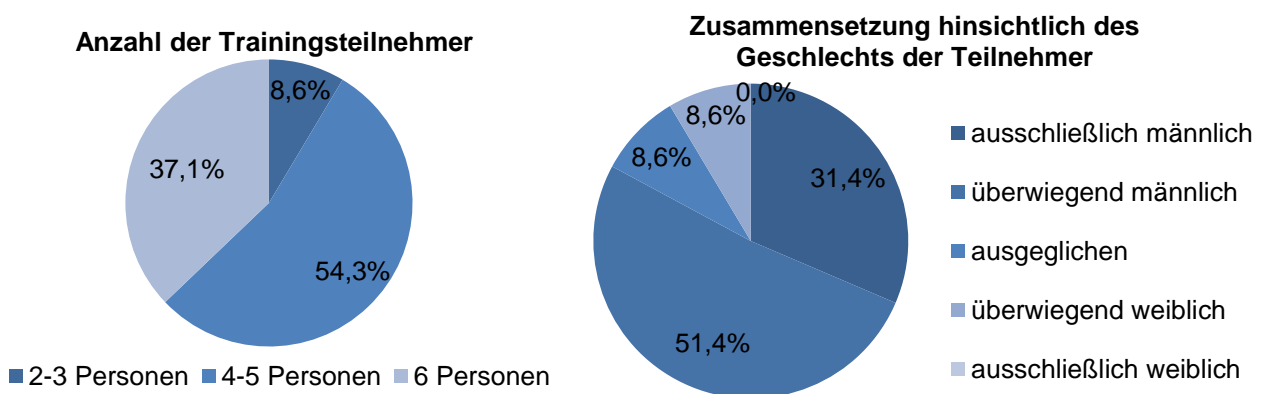


Abbildung 8: Zusammensetzung der durchgeführten Trainings

Nur knapp über ein Drittel der Trainings konnte wie geplant mit 6 Teilnehmern stattfinden. Dort konnten 71 Trainingsteilnehmer geschult werden. Weitere 88 hatten Gruppengrößen zwischen 4 und 5 Personen. Sieben Probanden konnten allerdings nur in sehr kleinen Gruppengrößen (2-3 Personen) geschult werden (insgesamt 3 Trainings). Fast ein Drittel der Trainings wurde nur mit männlichen Probanden durchgeführt

(insgesamt 51 Probanden). Weitere 93 Probanden (über die Hälfte der Trainings) hatten Trainings mit überwiegend männlichen Teilnehmern. Nur weibliche Teilnehmer kamen in keinem Training vor. Überwiegend weibliche Teilnehmer gab es in drei Trainings mit insgesamt 14 Probanden. 8 Probanden hatten ein ausgeglichenes Verhältnis zwischen männlichen und weiblichen Teilnehmern.

Hinsichtlich des Alters der Trainingsteilnehmer wiesen die meisten Trainings eine Mischung aus jungen (bis 30 Jahre) und mittelalten Teilnehmern (31-50 Jahre) auf (23 Trainings mit insgesamt 113 Probanden). Die zweitgrößte Gruppe waren Trainings mit einer heterogenen Zusammensetzung aller Altersklassen (8 Trainings mit 41 Probanden). Jeweils zwei Trainings gab es mit einer homogenen Alterszusammensetzung (2 Trainings mit 6 jungen und 2 Trainings mit 6 mittelalten Probanden). Die Rettungsdienst erfahrung war in den Trainings recht unterschiedlich verteilt. Am häufigsten kamen sehr gemischte Trainingszusammensetzungen vor, in denen alle Kategorien von Rettungsdienst erfahrung (Novizen mit unter 5 Jahren, Intermediates mit 5-10 Jahren und Experten mit über 10 Jahren Erfahrung; vgl. Abschnitt 6.5) vorlagen (22 Trainings mit 112 Probanden), gefolgt von Trainings mit einer Mischung von Novizen und Experten (8 Trainings mit 38 Probanden). Die Fahrer erfahrung erfasst mittels der Anzahl der Führerscheine war relativ gleichmäßig in den Trainings verteilt; erfasst in Dauer des Führerscheinbesitzes fanden sich bei den Gruppen alle Konstellationen von Erfahrungsgraden.

Zwischen den Trainings gibt es einen (projektimmanenten) systematischen Unterschied – alle sächsischen Teilnehmer wurden mit einem stationären Simulator geschult, während alle bayerischen Teilnehmer auf einem mobilen Simulator geschult wurden. Zudem stammte, der für alle Trainings identische Trainer, aus Sachsen. Vor diesem Hintergrund wurden t-Test für unverbundene Stichprobe hinsichtlich der unmittelbaren Bewertung des Trainings in den beiden Durchführungsorten durchgeführt – bei allen Aspekten (Trainingsinhalte, Trainer, Rahmenbedingungen, Anwendbarkeit, Weiterempfehlung und Gesamtnote) bewertete die sächsische Stichprobe das Training deutlich besser als die bayerische Stichprobe (alle T-Werte > 3.34; alle p = .000). Allerdings weisen massive Varianzinhomogenitäten<sup>2</sup> verursacht durch die bayerische Stichprobe darauf hin, dass hier die einzelnen Trainingsgruppen deutlich differenzieller bewertet wurden als in Sachsen (vgl. auch Abschnitt 4.2.1.1).

## 4.2 Hauptanalyse: Ergebnisse zu Effekten des Trainings und der Messungen

Tabelle 13 gibt einen deskriptiven Überblick über alle gemessenen abhängigen Variablen auf den verschiedenen Evaluationsebenen und deren Gesamtausprägung in der Stichprobe. Sie zeigt zudem den Streubereich der Variablen an, um eine Interpretation der Mittelwerte zu erlauben. Beispielsweise zeigt die Risikowahrnehmung bei schweren Regelverstößen einen sehr hohen Mittelwert bei sehr geringer Standardabweichung. Hier ist eine Erhöhung des Wertes schon aufgrund eines potenziellen Deckeneffektes (Bewertungen, die sich jetzt schon am oberen Ende der Bewertungsskala liegen) kaum möglich.

Tabelle 16: Überblick über die abhängigen Variablen zum Zeitpunkt T1 über alle Gruppen

Abhängige Variable	N	Möglicher Bereich	M	Min	Max	SD
Reaktionsebene - Trainingsbewertung						
Inhalte	166	[1; 5]	4.36	1.6	5.0	0.68
Trainer	166	[1; 5]	4.66	2.0	5.0	0.59
Rahmenbedingungen	166	[1; 5]	4.46	2.6	5.0	0.54

<sup>2</sup> Varianzinhomogenitäten zeichnen sich durch unterschiedliche Streuungen der Messwerte in den Teilstichproben aus. Die Probanden innerhalb der sächsischen Stichprobe haben das Training deutlich ähnlicher bewertet als die bayerische Stichprobe, in der die Probanden sehr unterschiedlich bewerteten.

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Möglicher Bereich	M	Min	Max	SD
Anwendbarkeit	166	[1; 5]	4.18	1.0	5.0	0.81
Weiterempfehlung	166	[1; 5]	4.54	1.0	5.0	0.91
Gesamtnote	166	[1; 6]	1.76	1.0	6.0	0.92
<b>Lernebene</b>						
Wissen	178	[1; 24]	8.67	2.5	17.8	2.60
Extraversion	179	[1; 4]	1.65	1.0	3.6	0.44
Sensation Seeking	179	[1; 4]	2.38	1.3	3.8	0.50
Gewissenhaftigkeit	178	[1; 4]	2.72	1.3	4.0	0.60
Unverträglichkeit	179	[1; 4]	2.32	1.0	3.6	0.52
Reaktanz	179	[1; 4]	1.51	1.0	3.6	0.45
Risikowahrnehmung bei schweren Regelverstößen	179	[1; 5]	4.44	3.4	5.0	0.33
Risikowahrnehmung bei mittelschweren Regelverstößen	179	[1; 5]	3.73	2.3	5.0	0.57
Risikowahrnehmung bei leichten Regelverstößen	179	[1; 5]	3.37	1.5	4.7	0.56
Risikowahrnehmung bei normalem Fahren	179	[1; 5]	2.78	1.5	4.5	0.59
Risk Seeking	178	[1; 5]	1.91	1.0	5.0	0.74
Fahrtskompetenz allgemein	178	[0; 10]	7.25	4.5	10.0	1.18
Fahrtskompetenz Rettungsdienst	178	[0; 10]	6.85	4.0	10.0	1.39
Unfallwahrscheinlichkeit allgemein	178	[0; 10]	4.74	1.0	10.0	2.05
Unfallwahrscheinlichkeit Rettungsdienst	178	[0; 10]	4.82	0.0	10.0	1.89
<b>Verhaltensebene</b>						
Durchschnittliche Geschwindigkeit (in km/h)	141	/	53.60	30.8	75.0	10.71
Maximale Geschwindigkeit (in km/h)	141	/	116.72	73.9	157.6	19.82
Maximale positive Längs-beschleunigung (in m/s <sup>2</sup> )	141	/	3.40	1.1	11.8	1.98
Mittlere positive Längs-beschleunigung (in m/s <sup>2</sup> )	141	/	0.32	0.2	1.0	0.14
Dauer Trigger positive Längs-beschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	141	/	0.03	0.0	1.3	0.16
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	141	/	0.01	0.0	0.8	0.09
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	141	/	-3.41	-1.3	-10.4	1.56
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	141	/	-0.32	-0.2	-1.0	0.13
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	141	/	0.93	0.0	27.0	3.40
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	141	/	0.32	0.0	10.9	1.35
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	141	/	0.02	0.0	1.5	0.15
Anzahl Trigger pro Minute negative Längsbeschleunigung (5,9 m/s <sup>2</sup> )	141	/	0.01	0.0	0.7	0.08
Maximale positive Querb beschleunigung (in m/s <sup>2</sup> )	141	/	5.04	2.0	22.0	3.42
Mittlere positive Querb beschleunigung (in m/s <sup>2</sup> )	141	/	0.48	0.2	1.7	0.29
Maximale negative Querb beschleunigung (in m/s <sup>2</sup> )	141	/	-4.61	-1.0	-20.6	3.14
Mittlere negative Querb beschleunigung (in m/s <sup>2</sup> )	141	/	-0.45	-0.2	-1.6	0.27
Dauer Trigger Querb beschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	141	/	0.77	0.0	23.1	3.13
Anzahl Trigger pro Minute Querb beschleunigung (6,9 m/s <sup>2</sup> )	141	/	0.33	0.0	8.6	1.31
Fahrerbeurteilung sichere Fahrweise	156	[0; 100]	81.28	0.0	100.0	12.92
Fahrerbeurteilung vorausschauende Fahrweise	156	[0; 100]	81.79	44.4	100.0	11.09



Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Möglicher Bereich	M	Min	Max	SD
Fahrerbeurteilung rasante Fahrweise	156	[0; 100]	39.40	0.0	94.4	20.96
Fahrerbeurteilung hektische Fahrweise	156	[0; 100]	16.23	0.0	72.2	14.02
Beifahrerbeurteilung sichere Fahrweise	147	[0; 100]	88.60	36.7	100.0	11.92
Beifahrerbeurteilung vorausschauende Fahrweise	147	[0; 100]	87.26	36.7	100.0	10.65
Beifahrerbeurteilung rasante Fahrweise	147	[0; 100]	34.34	0.0	100.0	21.77
Beifahrerbeurteilung hektische Fahrweise	147	[0; 100]	13.80	0.0	75.0	14.63
<b>Resultateebene</b>						
Kognitive Irritation	179	[1; 7]	2.78	1.0	6.0	1.26
Emotionale Irritation	179	[1; 7]	2.54	1.0	5.4	0.99
Irritation gesamt	179	[1; 7]	2.63	1.0	5.6	0.97
GBB Erschöpfung	176	[0; 24]	4.85	0.0	19.0	3.76
GBB Magenbeschwerden	179	[0; 24]	2.60	0.0	17.0	3.41
GBB Gliederschmerzen	179	[0; 24]	5.74	0.0	20.0	4.25
GBB Herzbeschwerden	178	[0; 24]	1.85	0.0	17.0	3.26
GBB Beschwerdedruck	175	[0; 96]	15.07	0.0	63.0	12.67
Wohlbefinden (WHO-5)	179	[0; 5]	3.22	1.0	5.0	0.89
Burnout (emotionale Erschöpfung)	179	[1; 6]	2.72	1.0	6.0	1.14
Burnout (Zynismus)	179	[1; 6]	2.59	1.0	5.3	1.13
Vitalität (BBS)	156	/	-0.86	-4.1	1.0	0.86
Gleichgewicht (BBS)	156	/	-0.33	-3.3	1.3	0.54
Extraversion (BBS)	156	/	-0.32	-2.8	1.5	0.71
Vigilanz (BBS)	156	/	-0.69	-4.1	1.1	0.75
NASA geistige Anforderungen	154	[1,100]	38.8	2.5	97.5	19.5
NASA körperliche Anforderungen	154	[1,100]	36.6	2.5	97.5	18.2
NASA zeitliche Anforderungen	154	[1,100]	43.7	1.5	97.5	19.4
NASA Leistung	153	[1,100]	81.9	18.5	99.2	11.7
NASA Anstrengung	153	[1,100]	35.4	0.7	97.5	18.5
NASA Frustration	154	[1,100]	22.1	0	94.3	19.2
Mittlere Müdigkeit (KSS)	154	/	3.49	1.0	6.7	1.23
Mittlere Ermüdung (KSS)	149	/	0.94	-3.5	6.3	1.84
Mittlere Herzfrequenz (Schläge pro Minute)	135	/	90.0	57	122	10.2
Maximale Herzfrequenz (Schläge pro Minute)	135	/	140.6	93	199	15.7
Minimale Herzfrequenz (Schläge pro Minute)	133	/	56.1	33	78	8.6
„Hilfsfrist“ (Dauer der Einsatzfahrten, in mm:ss)	150	/	06:41	02:39	14:14	02:16

Abkürzungen: N=Anzahl der gemessenen Probanden, M=Mittelwert, Min=Minimum, Max=Maximum, SD=Standardabweichung

In den folgenden Abschnitten werden die Trainingseffekte auf den verschiedenen Evaluationsebenen vorgestellt. Dabei wird die Reaktionsebene abweichend von den Analysen auf den weiteren Ebenen aufgrund der Unterschiede im Erhebungsvorgehen (siehe unten) zunächst vollständig vorgestellt, während für die weiteren Ebenen zunächst die Hauptbefunde im T1-T2-Vergleich vorgestellt werden, bevor in weiteren Kapiteln weitere Messzeitpunkte und Ergebnisse der nicht verbundenen Stichproben präsentiert werden.

## 4.2.1 Reaktionen auf die durchgeführten Trainings

Da aufgrund des Wartegruppen-Ansatzes alle Probanden das Training unabhängig von der Studiengruppe erhielten, werden diese Daten im Folgenden gesondert betrachtet. Neben Unterschieden zwischen den Versuchsgruppen werden qualitative Daten in Form der freien Kommentare zu den Trainings und die Veränderungen der Trainingsbewertungen im Zeitverlauf dargestellt.

### 4.2.1.1 Haupteffekte in der Reaktionsebene direkt nach dem Training

Wie bereits aus der vorherigen Tabelle 16 zu entnehmen, wird das Training insgesamt von allen Teilnehmer direkt nach dem Training sehr gut bewertet - auf einer Skala von 1 bis 5 wird kein Teilaspekt unter 4 eingeschätzt und die Gesamtbewertung via Schulnote liegt zwischen „gut“ und „sehr „gut“. Vor diesem Hintergrund werden in der folgenden Tabelle die Mittelwerte und Unterschiede zwischen den Studiengruppen hinsichtlich der Trainingsbewertungen betrachtet. Obwohl hier kein Unterschied erwartet und gewünscht wurde (die Gruppen unterscheiden sich nur in der Anzahl der Messungen vor dem Training), zeigten sich sehr deutliche Effekte: Die Wartegruppe (Gruppe 2) bewertet das Training durchgehend und hochsignifikant in allen Bereichen besser als die Interventionsgruppe (Gruppe 1). Mit einer minimalen Varianzaufklärung<sup>3</sup> von 7.2%, weisen die Unterschiede eine mittlere Effektstärke auf. Gruppe 3 bewertet das Training durchgehend zwischen Gruppe 1 und 2. Obwohl über die Gründe nur spekuliert werden kann, ist denkbar, dass hier Erwartungseffekte eine Rolle gespielt haben – mit der längeren Wartezeit auf das Training wurde möglicherweise die Bereitschaft zu einer besseren Bewertung erhöht, zumal Personen, denen das Training als „Anreiz“ zur Studienteilnahme zu diesem Zeitpunkt möglicherweise bereits aus der Studie ausgestiegen waren (vgl. Abbildung 4).

Tabelle 17: Haupteffekte der Evaluationsebene „Reaktion“

Evaluierte Skalen	Gruppe 1 (a) (N=73)		Gruppe 2 (b) (N=68)		Gruppe 3 (c) (N=25)		Gruppeneffekt	
	M	SE	M	SE	M	SE	F	p
Inhalte	4.15 <sup>b</sup>	0.07	4.57 <sup>a</sup>	0.08	4.37	0.13	<b>7.32</b>	.001***
Trainer	4.48 <sup>b</sup>	0.07	4.83 <sup>a</sup>	0.07	4.73	0.11	<b>6.95</b>	.001***
Rahmenbedingungen	4.28 <sup>b,c</sup>	0.06	4.61 <sup>a</sup>	0.06	4.55 <sup>a</sup>	0.10	<b>8.11</b>	.000***
Anwendbarkeit	3.95 <sup>b</sup>	0.09	4.42 <sup>a</sup>	0.09	4.25	0.15	<b>6.65</b>	.002**
Empfehlung	4.28 <sup>b</sup>	0.10	4.79 <sup>a</sup>	0.11	4.63	0.17	<b>6.28</b>	.002**
Gesamtnote	2.02 <sup>b</sup>	0.10	1.46 <sup>a</sup>	0.11	1.85	0.18	<b>7.04</b>	.001***

Berechnung: Varianzanalyse unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Gruppen

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Zusätzlich zeigte sich ein deutlicher Einfluss des Alters auf die Trainingsbewertung. Insgesamt bewerten die älteren Probanden (Mediansplit; ab 32 Jahren) das Training besser (vgl. Tabelle 18). In den Gruppen 1 und 2 trifft das auf alle Variablen zu. In Gruppe 3 bewerten die jüngeren das Training immer besser als die älteren Probanden. Abbildung 9 zeigt beispielhaft die unterschiedliche Bewertung des Trainings anhand der Inhalte des Trainings für die drei Versuchsgruppen in Abhängigkeit des Alters der Probanden.

<sup>3</sup> Die Varianzaufklärung gibt an, wie gut das berechnete Modell die tatsächliche Streuung der Daten erklären kann. Mithilfe der Varianzaufklärung kann die Effektstärke ermittelt werden, die die praktische Relevanz der Ergebnisse unabhängig der Stichprobengröße angibt.

Tabelle 18: Unterschiede in der Trainingsbewertung in Abhängigkeit vom Alter

Evaluierte Skalen	Jüngere Probanden		Ältere Probanden		Gruppenvergleich	
	M	SD	M	SD	t	p
Inhalte	4.20	0.733	4.51	0.581	-3.00	.003**
Trainer	4.56	0.674	4.76	0.479	-2.23	.027*
Rahmenbedingungen	4.35	0.553	4.56	0.509	-2.51	.013*
Anwendbarkeit	4.03	0.863	4.33	0.727	-2.45	.015*
Empfehlung	4.35	1.058	4.73	0.683	-2.69	.008**
Gesamtnote	1.92	0.970	1.61	0.856	-2.19	.030*

t-Test für unverbundene Stichproben, korrigiert für Varianzheterogenität; p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Altersmedianschnitt=32 Jahre

Auch diese Effekte waren nicht erwartet und können nur spekulativ interpretiert werden. Denkbar ist, dass insbesondere die bereits etwas ältere „Anmutung“ des Simulators in der jüngeren Gruppe, die vermutlich mit neueren Generationen von Simulatoren, die auf jedem PC laufen, vertrauter ist, zu schlechteren Bewertungen in dieser Gruppe geführt haben könnten (vgl. auch Qualitative Daten). Auch Ähnlichkeit mit dem Trainer könnte (als sozialpsychologischer Effekt) hier zu einer höheren Sympathie in der älteren Gruppe geführt haben. Die differenziellen Effekte hinsichtlich Gruppe 3 sind hiermit nicht zu erklären, allerdings sind diese Daten aufgrund der sehr kleinen Gruppengröße hier eher vorsichtig zu betrachten.

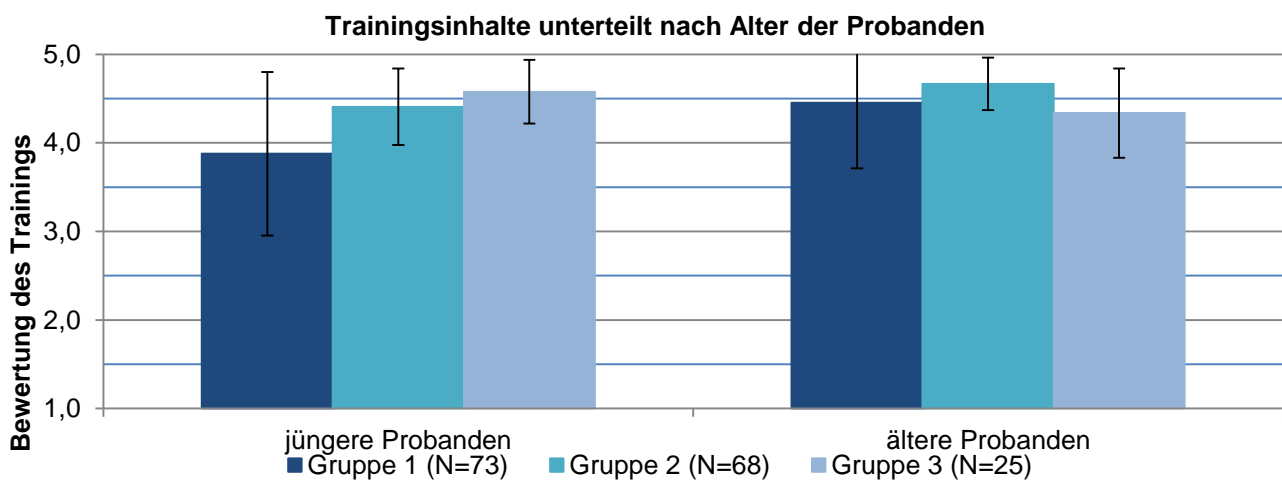


Abbildung 9: Bewertung der Trainingsinhalte in Abhängigkeit von Versuchsgruppe und Alters der Probanden

Wie bereits in Abschnitt 4.1 dargestellt, gab es hinsichtlich des Ortes systematische Unterschiede, die sich auch in den Bewertungen des Trainings ausdrücken. Ohne Kontrolle, aber auch unter Kontrolle von Alter und Geschlecht, zeigen sich deutliche Unterschiede zwischen den Standorten. Alle evaluierten Skalen wurden in Sachsen besser bewertet als in Bayern. Wie aufgrund der Aufteilung der Standorte auf die Studiengruppen zu erwarten, sind diese Unterschiede - wie aus den Interaktionseffekten erkennbar - relativ gleichmäßig auf die Gruppen verteilt. Lediglich bei der Bewertung des Trainers und der Gesamtnote gibt es einen tendenziellen Interaktionseffekt. Der Trainer wird in Sachsen immer besser bewertet als in Bayern, in der Gruppe 1 ist der Unterschied jedoch höher als in den anderen beiden Gruppen. Die Gesamtnote ist in allen drei Versuchsgruppen in Sachsen besser als in Bayern, der Unterschied ist in der Gruppe 2 jedoch geringer als in den anderen beiden Gruppen.

Tabelle 19: Unterschiede in den Reaktionen zwischen den Orten der Trainingsdurchführung

Evaluierte Skalen	Sachsen (N=58)		Bayern (N=108)		Haupteffekt Ort		Interaktionseffekt Gruppe*Ort	
	M	SD	M	SD	F	p	F	p
Inhalte	4.61	0.40	4.22	0.75	<b>11.26</b>	.001***	1.64	.197
Trainer	4.83	0.28	4.57	0.69	3.71	.056 <sup>+</sup>	2.55	.081 <sup>+</sup>
Rahmenbedingungen	4.69	0.35	4.33	0.58	<b>12.16</b>	.001***	0.81	.448
Anwendbarkeit	4.49	0.50	4.02	0.89	8.54	.004**	1.80	.169
Empfehlung	4.83	0.38	4.39	1.06	7.16	.008**	2.28	.105
Gesamtnote	1.40	0.53	1.95	1.03	11.24	.001***	2.67	.072 <sup>+</sup>

Berechnung: Varianzanalyse unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001;

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Anhand der Gesamtnote kann der Unterschied zwischen den Standorten noch einmal deutlich gemacht werden: In Sachsen wurde insgesamt nur einmal die Note 3 vergeben, alle anderen Probanden haben dem Training eine 1 oder 2 gegeben. In Bayern wurde das Training teilweise deutlich schlechter bewertet. Insgesamt einmal wurde die Note 6 vergeben (Gruppe 1), fünfmal die Note 5 (alle Gruppe 1) und dreimal die Note 4 (zweimal Gruppe 1, einmal Gruppe 3). Die drei Vieren wurden in drei verschiedenen Trainings vergeben. Die fünf und die sechs wurden hingegen alle innerhalb einer Trainingsgruppe relativ zu Beginn des Projekts vergeben. In diesem Training schien der Ablauf für die Teilnehmer nicht den Erwartungen entsprochen zu haben und die negative Stimmung hat sich vermutlich hochgeschaukelt. Vor diesem Hintergrund wurde vom Projektteam eine Änderung in den Akquiseaktivitäten, vor allem in den Beschreibungen und Einladungen zu den Trainings, vorgenommen, um den potenziellen Teilnehmern realistischere Vorstellungen von dem Training zu geben.

Da es eine gleichmäßige Verteilung der Standorte auf die Studiengruppen und kaum Interaktionseffekte zwischen den Gruppen und dem Ort gibt, wird in den Analysen weiterhin lediglich das Alter und Geschlecht der Probanden durchgehend kontrolliert.

#### 4.2.1.2 Reaktionen auf das Training im langfristigen Verlauf

In diesem Abschnitt werden die Bewertungsverläufe für die Bereiche, die mehrfach erfasst wurden (Anwendbarkeit, Empfehlung und Gesamtnote) präsentiert. In einem ersten Schritt werden dabei die Bewertungen direkt nach dem Training sowie 1-2 Monate nach dem Training betrachtet, in einem weiteren Schritt dann die längerfristigen Bewertungen aus den Follow-up-Messungen. Eine Weile nach dem Training wird dieses in allen erfassten Bereichen schlechter bewertet als direkt danach. Zusätzlich gibt es Unterschiede zwischen den Gruppen, die über die Zeit gleich bleiben. Die Gruppe 2 bewertet das Training insgesamt besser als die anderen beiden Gruppen. Da es keinen Interaktionseffekt zwischen der Zeit und den Gruppen gibt, verläuft die Abnahme der Trainingsbewertung in allen Gruppen ähnlich.

Zur Bewertung des längerfristigen Erfolges werden nun die Follow-up-Messungen betrachtet, die sechs Monate nach dem Training durchgeführt wurden. Da die Zahl der Follow-up-Messungen deutlich geringer ist, als die 1-2 Monate nach dem Training, werden die Unterschiede deskriptiv dargestellt. Tabelle 21 zeigt die Mittelwerte der Bewertungen über die Probanden, die zu allen Zeitpunkten eine Trainingsevaluation ausgefüllt haben. Die Abnahme der Trainingsbewertung, die bereits 1-2 Monate nach dem Training gefunden werden kann, bleibt 6 Monate nach dem Training stabil.

Tabelle 20: Zeitlicher Verlauf in der Trainingsbewertung direkt nach dem Training sowie 1-2 Monate später

Abhängige Variable	Zeitpunkt	Gruppe 1 (N=64)		Gruppe 2 (N=64)		Gruppe 3 (N=23)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
		M	SD	M	SD	M	SE	F <sub>(1,146)</sub>	p	F <sub>(2,146)</sub>	P	F <sub>(2,146)</sub>	p
Anwendbarkeit	direkt	4.07	0.09	4.39	0.09	4.35	0.15	4.12	.044*	3.85	.023*	0.29	.747
	später	3.38	0.12	3.60	0.12	3.56	0.20						
Empfehlung	direkt	4.44	0.10	4.75	0.10	4.70	0.17	6.33	.013*	2.50	.085 <sup>+</sup>	0.26	.772
	später	3.68	0.15	3.88	0.15	3.78	0.25						
Gesamtnote	direkt	1.86	0.10	1.49	0.10	1.80	0.18	7.30	.008**	5.10	.007**	0.98	.376
	später	2.64 <sup>b</sup>	0.13	2.24 <sup>a</sup>	0.13	2.30	0.21						

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

Das Alter wirkt sich in der Empfehlung weiterhin wie bisher aus (F<sub>(1,39)</sub>=4.66, p=.037), die älteren bewerten das Training insgesamt etwas besser als ihre jüngeren Kollegen. In der Note gibt es keinen Gesamtalterseinfluss mehr, aber eine Interaktion zwischen dem Alter und den Zeitpunkten. Die Note sinkt im zeitlichen Verlauf bei den jüngeren Probanden stärker als bei den älteren (F<sub>(2,78)</sub>=3.47, p=.036). Der Einfluss des Erhebungsortes verhält sich wie in den bisherigen Berechnungen. Die sächsische Bewertung bleibt insgesamt besser als die bayerische. Die Abnahmen sind in beiden Gruppen vergleichbar.

Tabelle 21: Langfristiger zeitlicher Verlauf in der Trainingsbewertung direkt nach dem Training, 1-2 Monate später sowie 6 Monate später

Abhängige Variable	Zeitpunkt	Gruppe 1 (N=24)		Gruppe 2 (N=12)		Gruppe 3 (N=8)	
		M	SD	M	SD	M	SE
Anwendbarkeit	direkt	4.11	0.96	4.33	0.74	4.18	0.84
	Später (1-2 Monate)	3.64	1.11	3.66	0.90	3.08	0.91
	Follow-up (6 Monate)	3.66	0.95	3.43	0.83	3.18	0.89
Empfehlung	direkt	4.46	1.06	4.75	0.45	4.38	0.74
	Später (1-2 Monate)	4.02	1.20	4.25	0.97	3.38	0.92
	Follow-up (6 Monate)	3.92	1.25	4.25	0.75	3.25	1.16
Gesamtnote	direkt	1.96	1.27	1.38	0.48	2.13	0.99
	Später (1-2 Monate)	2.50	1.29	2.21	0.50	2.75	1.04
	Follow-up (6 Monate)	2.40	1.22	2.25	0.75	2.75	0.89

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

### 4.2.1.3 Freie Kommentare zum Training

Neben den erfassten Skalen wurden die Teilnehmer auch um freie Kommentare zum Training gebeten. Dabei wurden getrennte Fragen gestellt nach dem, was die Teilnehmer gut fanden und nach dem, was sie am Training nicht so gut fanden oder welche Verbesserungsvorschläge sie haben. Bei den 166 abgegebenen Evaluationsbögen wurden von 107 Probanden positive Kommentare zum Training angegeben. Insgesamt

wurden dabei 167 positive Punkte angesprochen, die zu Oberbegriffen zusammengefasst wurden. In Abbildung 10 sind alle Punkte unterteilt nach den Versuchsgruppen dargestellt.

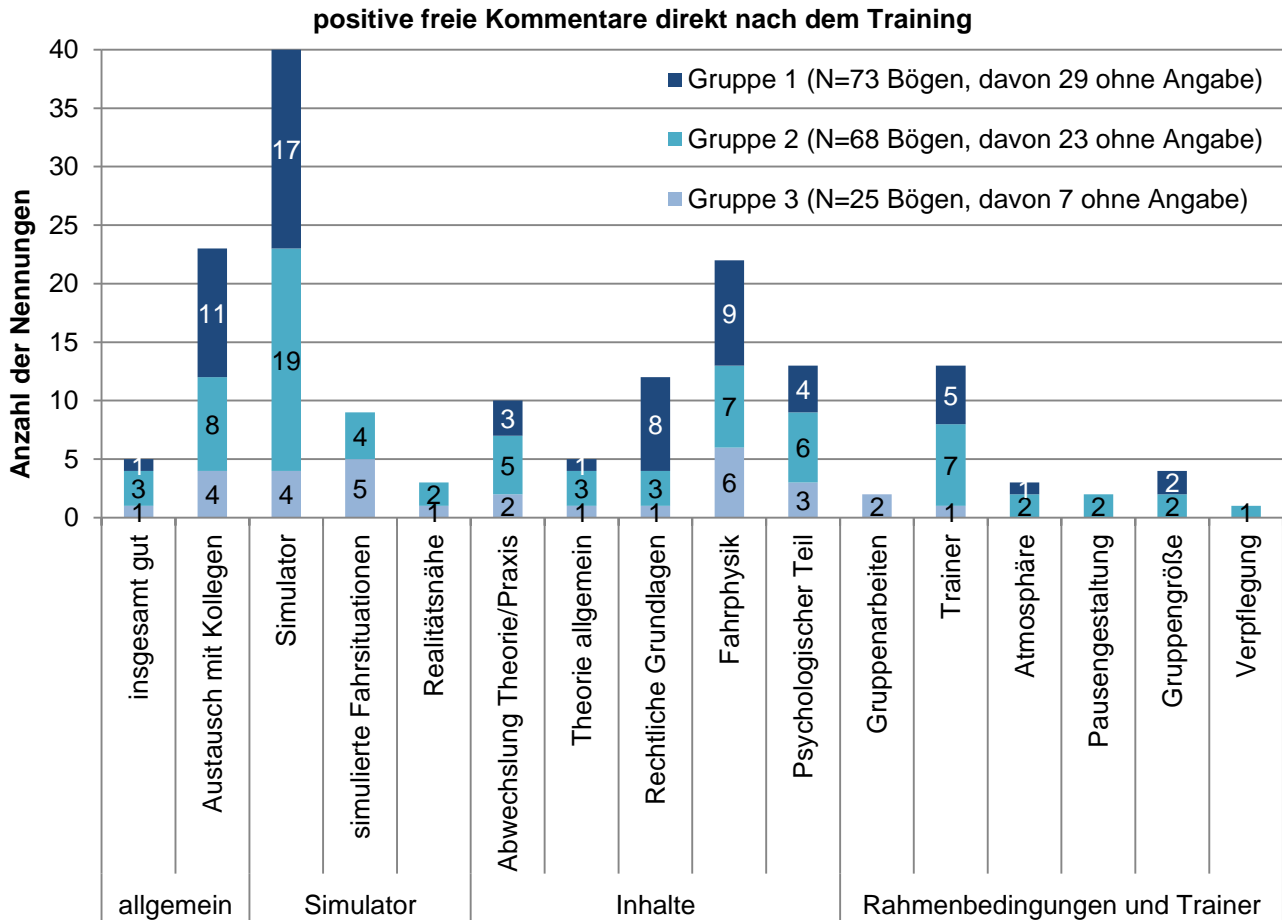


Abbildung 10: positive Kommentare direkt nach dem Training in Abhängigkeit von der Versuchsgruppe

Am häufigsten, von insgesamt 40 Probanden, wurden der Simulator und die Praxisteile im Training als positiv genannt, gefolgt von dem Austausch mit Kollegen und das gegenseitige Feedback (insgesamt 23 Nennungen). In den theoretischen Inhalten wurden Themen der Fahrphysik mit Einflüssen der Geschwindigkeit und Bremswegen am häufigsten genannt als das, was die Probanden für sich mitnehmen (22 Nennungen). Auch die psychologischen Aspekte des Fahrens (13 Nennungen) und die rechtlichen Grundlagen (12 Nennungen) wurden häufig als positive Punkte angegeben. Der Trainer wurde insgesamt 13-mal gelobt. Die weiteren Nennungen wurden seltener genannt, immerhin noch zehn Probanden fanden die Abwechslung zwischen theoretischen und praktischen Blöcken gut, neun die Fahrsituationen im Simulator. Nur sehr wenige nannten organisatorische Dinge, wie die Gruppengröße oder Verpflegung. Ein paar Beispiele für Kommentaren über Inhalte, die die Probanden hilfreich fanden, sind: „Sensibilisierung auf eigenes Verhalten (bzw. Fehlverhalten)“, „Abwechslung Theorie + praktisches Simulationstraining; Verdeutlichung der Gefahren bei erhöhter Geschwindigkeit“ oder „Ich muss erst sacken lassen. Auf jeden Fall gehört so ein Training in jede Ausbildung für Bereiche FW, Pol, THW, u. RD mit hinein.“

Bei der Frage, nach eher negativen Aspekten oder Verbesserungsvorschlägen, gaben 138 Personen nichts an. Von den 28 Antwortenden wurde am häufigsten die zu lange Theorie genannt (13 Nennungen) gefolgt von der zu kurzen Zeit im Simulator (sechs Nennungen). Der Simulator war fünf Personen zu unrealistisch. Alle anderen Nennungen waren seltener. Explizite theoretische Inhalte, die zu viel waren, wurden nur

hinsichtlich der rechtlichen Grundlagen genannt. Sehr negativ formulierte Bewertungen fanden sich nur sehr selten, z.B. „leider war nichts hilfreich“, „sehr viel Theorie, viel Reden d. Ausbilder“, „dauernde Missverständnisse zw. Ausbilder & TN“ oder „ca 98% des Trainings war nicht praxisorientiert bzw. nutzbar im Einsatz“.

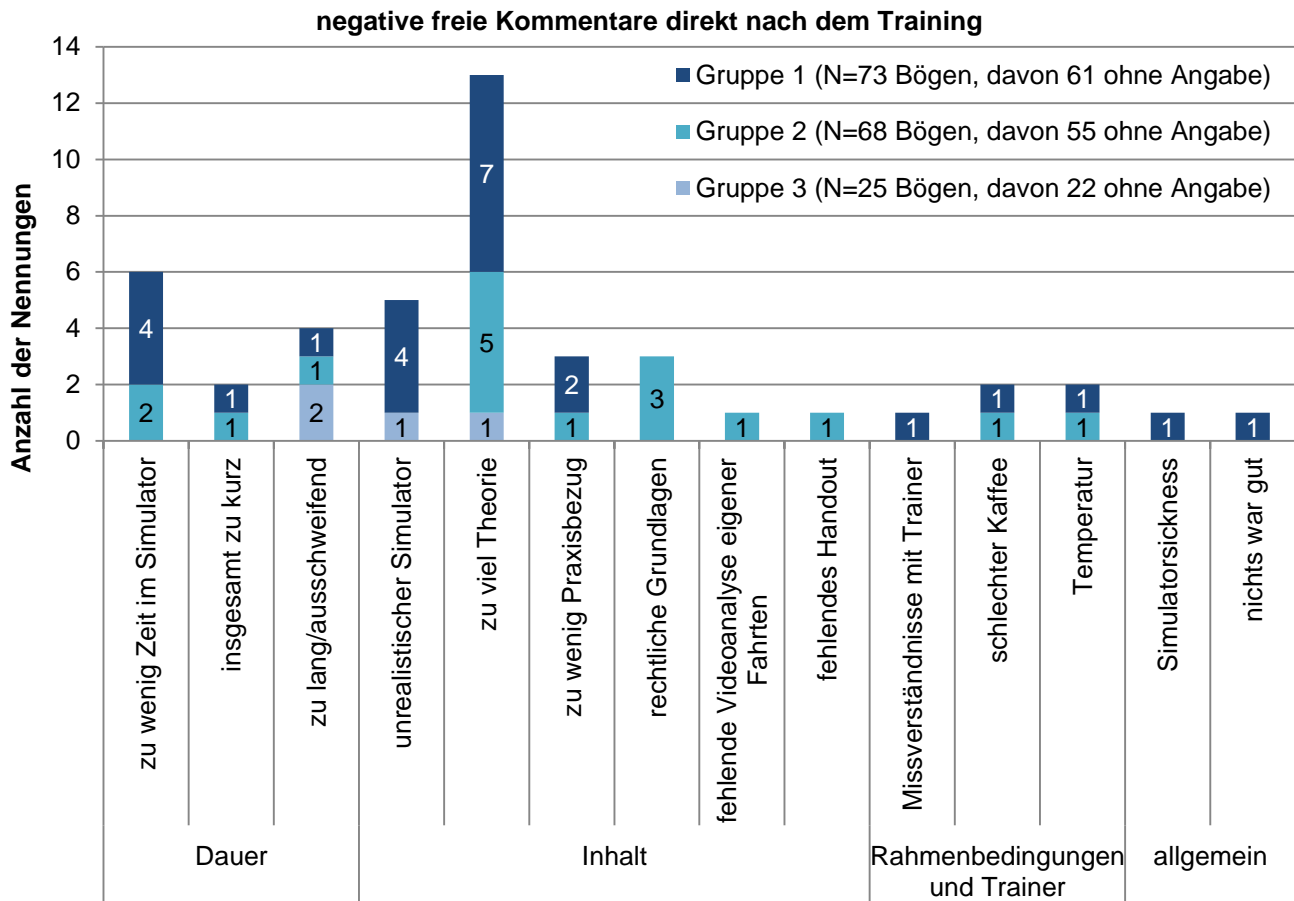


Abbildung 11: negative Kommentare direkt nach dem Training in Abhängigkeit von der Versuchsgruppe

Neben diesen unmittelbaren Reaktionen wurden die Teilnehmer in den folgenden Fragebögen jeweils nur noch gefragt, ob es etwas gibt, was sie uns im Nachhinein zum Training noch mitteilen möchten. In den 158 Bögen, die 1-2 Monate nach dem Training ausgefüllt wurden, gab es von 59 Probanden Kommentare mit insgesamt 98 Aspekten. Positive Nennungen gab es insgesamt nur 14-mal (die ersten vier Kategorien der Abbildung 12). Explizite Verbesserungsvorschläge wurden insgesamt am häufigsten genannt (von Kategorie „regelmäßige Wiederholungen wären sinnvoll“ bis Kategorie „Navi im Simulator lenkt ab“). Die letzten fünf Kategorien zeigten eher negative Kommentare zum Training oder zur Organisation.

Insgesamt am häufigsten negativ wurde die zu kurze Zeit im Simulator genannt (25 Nennungen) gefolgt davon, dass die Probanden den Simulator als zu unrealistisch, veraltet oder als praxisfremd wahrgenommen haben (16 Nennungen). Acht Probanden gaben an, dass ihnen ein Fahrsicherheitstraining mehr gebracht hätte oder zumindest zusätzlich gut wäre. Sieben würden das vorliegende Training eher für unerfahrene Einsatzkräften empfehlen. Verbesserungsvorschläge wurden zum Beispiel wie folgt gegeben: „Das Training im Fahrsimulator war teilweise ein wenig unrealistisch, z.B. das entgegenkommende Fahrzeuge nicht dem Sondersignalfahrzeug ausweichen.“, „Zu kurzer Zeitraum. Bitte mehr Simulationsfahrten! Training ist sinnvoll VOR der ersten Blaulichtfahrt, nicht wenn man schon lange Blaulicht fährt.“ Oder „Es wäre interessant sich nach dem Simulator Training selbst einmal fahren zu sehen (Wieder-

holung).“. Aber auch die negativen Kommentare wurden teilweise sehr genau begründet: „[...] Meine Vorstellungen/Erwartungen sind überhaupt nicht erfüllt worden, man lernt nicht, wenn man 2x2min am Fahrsimulator sitzt und irgendwelche künstlichen "Stressfaktoren" herbeigeführt werden ("von 180 in 3-Schritten rückwärts zählen"). Ich würde mir ein reales Fahrtraining wünschen und nicht an einem Computer. Wenn am PC dann unter realen Bedingungen. Schade, hatte mir mehr erhofft.“

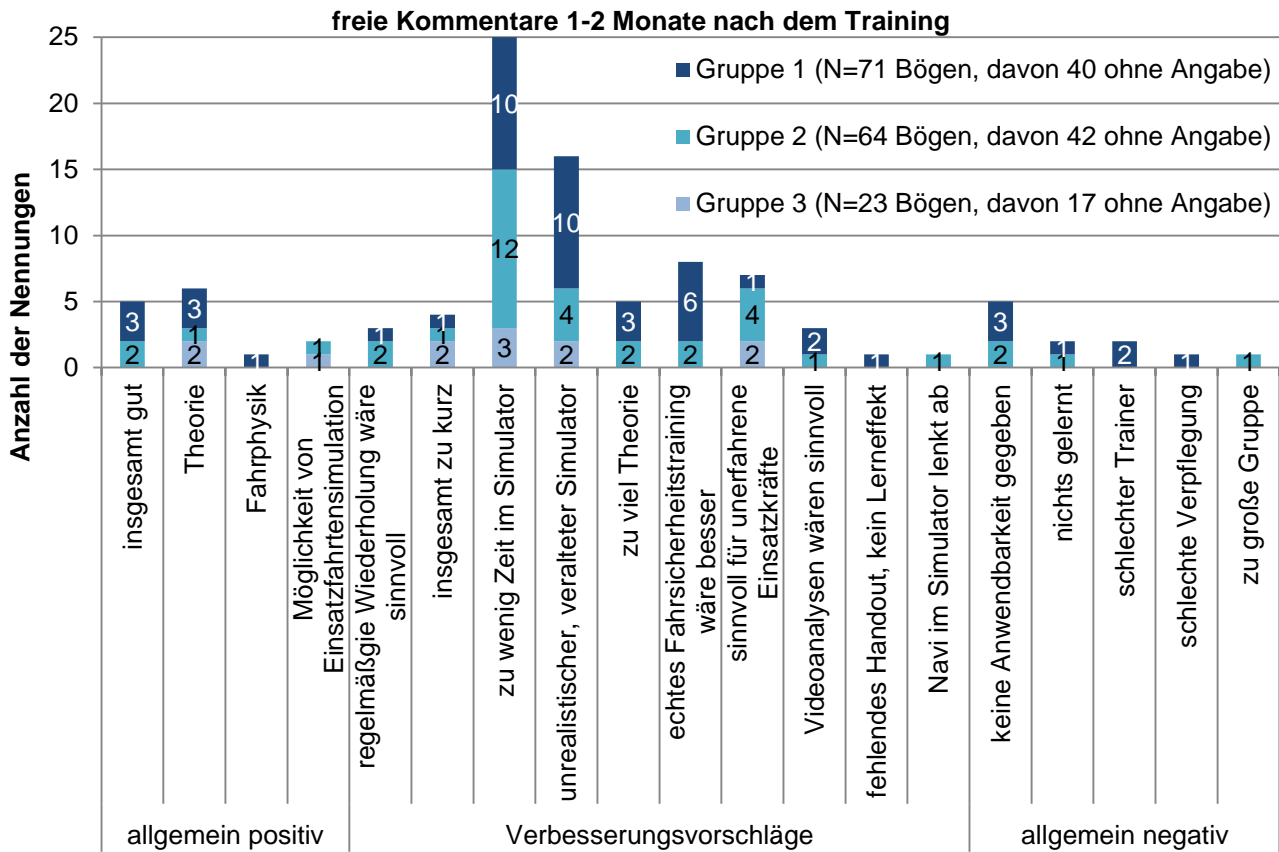


Abbildung 12: Freie Antworten 1-2 Monate nach dem Training in Abhängigkeit von der Versuchsgruppe

In den Follow-up-Messungen machten 14 Personen bei 45 Antwortbögen noch weitere freie Angaben. Die insgesamt genannten 20 Aspekte (vgl. Abbildung 13) ähnelten den Aussagen, die 1-2 Monate nach dem Training gemacht wurden. Am häufigsten wurde der Simulator kritisiert (sieben Nennungen) gefolgt davon, dass das Training eher für unerfahrene Einsatzkräfte als sinnvoll erachtet wird (vier Nennungen). Zu wenig Zeit im Simulator wurde noch zweimal genannt, alle andere Punkte nur einmal. Die Kommentare wurden insgesamt kürzer, dennoch haben einige Probanden sich auch bei der letzten Rückmeldung die Mühe gemacht, längere Erklärungen zu schreiben, zum Beispiel: „Der Fahrsimulatorteil ist der Schwachpunkt: zu wenige Einheiten; nicht realistisch genug (bremsen/lenken/beschleunigen)“ oder „Das Training ist sinnvoll für Berufseinsteiger um häufige Gefahrensituationen bei Alarmfahrten kennenzulernen. Für erfahrene Fahrer im Rettungsdienst ist das Training irrelevant.“

#### 4.2.2 Vergleich aller drei Studiengruppen zu den Zeitpunkten T1 und T2

In den folgenden Abschnitten werden für die weiteren drei Evaluationsebenen (Lernen, Verhalten, Resultate) jeweils zunächst die Trainingseffekte für die verschiedenen betrachteten Variablen im „belastbarsten“ Design, dem Interventions-Kontrollgruppen-Vergleich zwischen erstem und zweitem Messzeitpunkt, dargestellt, um zu ermitteln, ob sich die Gruppen mit Trainingsteilnahme im Verlauf von denen



ohne Training unterscheiden. Es werden jeweils zunächst alters- und geschlechtskontrollierte zweifaktorielle Varianzanalysen mit Messwiederholung dargestellt. Im Anschluss werden die Effekte genauer erläutert und Einflussvariablen beschrieben.

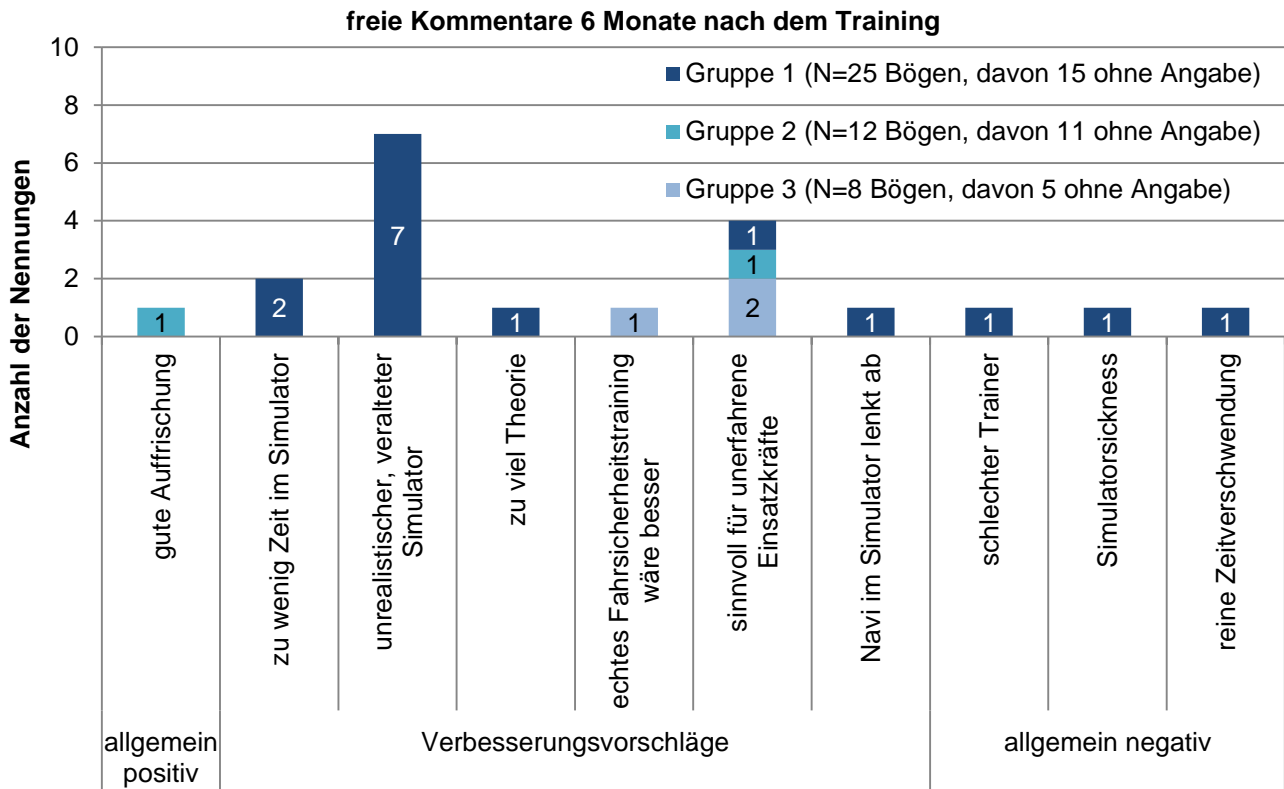


Abbildung 13: Freie Antworten 6 Monate nach dem Training in Abhängigkeit von der Versuchsgruppe

#### 4.2.2.1 Haupteffekte in der Lernebene

Tabelle 22 gibt einen Überblick über alle in der Lernebene erfassten abhängigen Variablen. Sie zeigt die Mittelwerte in den drei Versuchsgruppen sowie die Haupteffekte und den Interaktionseffekt für den zeitlichen Verlauf und zwischen den Gruppen. Im Weiteren werden die einzelnen Effekte genauer beschrieben.

Insgesamt finden sich nur vergleichsweise wenige Interaktionseffekte, die auf die Trainingswirksamkeit hinweisen. Diese liegen im Bereich Wissen, Extraversion, Risikowahrnehmung bei Normalfahrten und subjektiv wahrgenommene Fahrkompetenz. Im Folgenden werden die einzelnen Punkte näher dargestellt.

*Wissen:* Der hochsignifikante Interaktionseffekt für die Veränderungen in den Wissenstestwerten der verschiedenen Studiengruppen ist der einzige der aufgefunden Effekte, der eine mittlere Effektstärke aufweist. Die Varianzaufklärung durch den Interaktionseffekt beträgt 8%. Abbildung 14 zeigt graphisch den Verlauf in der Gesamtpunktzahl Wissen für die drei Versuchsgruppen.

Tabelle 22: Haupteffekte der Evaluationsebene „Lernen“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Gruppe 3 (c)		Zeiteffekt		Gruppeneffekt		Interaktionseffekt	
			M	SE	M	SE	M	SE	F	p	F	p	F	p
Wissen														
Wissen	170	T1	8.31	0.30	9.03	0.30	9.03	0.51	0.65	.420	0.46	.635	<b>7.16</b>	.001** *
		T2	9.99	0.31	8.77	0.30	9.56	0.52						
Verkehrsrelevante Persönlichkeitseigenschaften														
Extraversion	164	T1	1.66	0.05	1.66	0.05	1.54	0.08	0.45	.503	0.13	.877	3.01	.052 <sup>+</sup>
		T2	1.60	0.05	1.63	0.05	1.66	0.08						
Sensation Seeking	164	T1	2.30	0.06	2.38	0.06	2.38	0.10	0.52	.470	0.43	.650	0.32	.726
		T2	2.30	0.06	2.35	0.05	2.31	0.09						
Gewissenhaftigkeit	163	T1	2.74	0.07	2.75	0.07	2.52	0.12	1.62	.205	1.15	.320	0.71	.494
		T2	2.65	0.07	2.69	0.06	2.56	0.11						
Unverträglichkeit	164	T1	2.30	0.07	2.33	0.06	2.32	0.10	0.00	.994	0.13	.878	0.16	.849
		T2	2.21	0.06	2.27	0.05	2.27	0.09						
Reaktanz	164	T1	1.43	0.05	1.53	0.05	1.49	0.08	1.05	.307	0.99	.375	0.21	.811
		T2	1.45	0.05	1.53	0.05	1.53	0.08						
Risikobewusstsein														
Risikowahrnehmung bei schweren Regelverstößen	164	T1	4.44	0.04	4.44	0.04	4.52	0.07	1.59	.209	0.56	.572	0.37	.690
		T2	4.50	0.04	4.45	0.04	4.52	0.07						
Risikowahrnehmung bei mittelschweren Regelverstößen	164	T1	3.79	0.06	3.65	0.06	3.80	0.10	1.19	.277	2.55	.081 <sup>+</sup>	0.67	.513
		T2	3.94 <sup>b</sup>	0.07	3.72 <sup>a</sup>	0.06	3.95	0.11						
Risikowahrnehmung bei leichten Regelverstößen	164	T1	3.43	0.07	3.30	0.06	3.47	0.11	0.24	.622	3.57	.030*	1.64	.198
		T2	3.54 <sup>b</sup>	0.07	3.29 <sup>a</sup>	0.06	3.60	0.11						
Risikowahrnehmung bei normalem Fahren	164	T1	2.81	0.07	2.75	0.06	2.85	0.11	0.11	.736	2.30	.103	4.10	.018*
		T2	2.96 <sup>(b)</sup>	0.07	2.73 <sup>(a)</sup>	0.06	3.06	0.10						
Risk Seeking	163	T1	1.83	0.09	1.84	0.09	2.13	0.15	1.42	.236	3.12	.047*	1.43	.243
		T2	1.81 <sup>c</sup>	0.09	1.66 <sup>c</sup>	0.08	2.12 <sup>a,b</sup>	0.14						

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Gruppe 3 (c)		Zeiteffekt		Gruppeneffekt		Interaktionseffekt	
			M	SE	M	SE	M	SE	F	p	F	p	F	p
Subjektive Fahrfähigkeit														
Fahrtkompetenz allgemein	163	T1	6.87	0.14	7.43	0.13	7.69	0.23	0.03	.861	4.03	.020*	3.91	.022*
		T2	7.11 <sup>b,c</sup>	0.13	7.31 <sup>a,(c)</sup>	0.12	7.42 <sup>a,(b)</sup>	0.21						
Fahrtkompetenz Rettungsdienst	163	T1	6.50	0.17	7.02	0.16	7.28	0.27	0.19	.666	2.51	.084 <sup>+</sup>	2.21	.113
		T2	6.69 <sup>c</sup>	0.18	6.77 <sup>(c)</sup>	0.17	7.18 <sup>(b),c</sup>	0.29						
Unfallwahrscheinlichkeit allgemein	163	T1	4.47	0.26	5.01	0.24	4.79	0.41	0.80	.372	2.16	.119	0.54	.584
		T2	4.66 <sup>b</sup>	0.23	5.20 <sup>a</sup>	0.21	4.47	0.36						
Unfallwahrscheinlichkeit Rettungsdienst	163	T1	4.66	0.24	4.97	0.22	4.96	0.38	5.88	.016*	3.09	.048*	1.41	.246
		T2	4.32 <sup>b,(c)</sup>	0.24	5.19 <sup>a</sup>	0.23	5.36 <sup>a</sup>	0.39						

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: N=Anzahl der Probanden, M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Gruppen

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Gruppe 1 startet von einem geringeren Wissensniveau als die anderen Gruppen, zeigt dann aber die größten Zuwächse. Auch Gruppe 3, als Interventionsgruppe ohne Messungen vor dem Training, weist einen Zuwachs auf, während die Wartegruppe ohne Training ein leicht abfallendes Niveau hat. Damit sind die Änderungen im Wissen hypothesenkonform. Auffällig ist allerdings auch das allgemeine Niveau der erzielten Testwerte. So konnten, selbst nach dem Training von der Gruppe mit dem höchsten Mittelwert (Gruppe 1) im Mittel nicht einmal die Hälfte der möglichen Punkte erreicht werden. Im Weiteren sollen die einzelnen Wissenstestbereiche genauer betrachtet werden, um herauszufinden, in welchem Bereich der Wissenszuwachs lag (vgl. Tabelle 23).

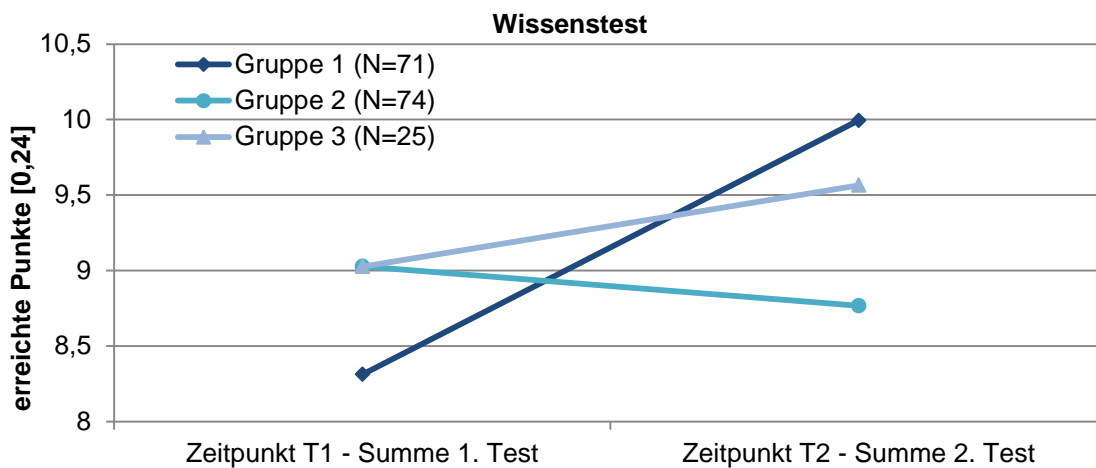


Abbildung 14: Erreichte Gesamtpunkte im Wissenstest im zeitlichen Verlauf

Tabelle 23: Unterschiede in den verschiedenen Wissenstestbereichen

Wissensbereiche	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Gruppe 3 (c)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
		M	SE	M	SE	M	SE	F <sub>(1,165)</sub>	p	F <sub>(2,165)</sub>	p	F <sub>(2,165)</sub>	p
Unfallursachen & -situationen	T1	1.59	0.13	1.74	0.13	1.49	0.22	0.04	.840	0.86	.423	4.51	.012*
	T2	2.35	0.15	1.81	0.15	2.31	0.25						
Rechtliche Grundlagen	T1	2.57	0.13	2.39	0.13	2.69	0.22	0.56	.456	1.56	.213	0.64	.529
	T2	2.73 <sup>(b)</sup>	0.13	2.43 <sup>(a)</sup>	0.13	2.49	0.22						
Fahrphysik	T1	1.56	0.13	2.09	0.13	1.90	0.23	0.10	.756	0.74	.479	4.02	.020*
	T2	1.82	0.14	1.65	0.14	1.73	0.23						
Informationsverarbeitung	T1	0.49	0.08	0.46	0.08	0.53	0.13	0.05	.822	3.28	.040*	3.54	.031*
	T2	0.89 <sup>b</sup>	0.08	0.48 <sup>a</sup>	0.08	0.73	0.14						
Entscheidungsverhalten	T1	1.60	0.06	1.68	0.06	1.73	0.11	0.05	.820	1.51	.224	1.39	.253
	T2	1.66 <sup>(c)</sup>	0.06	1.62 <sup>(c)</sup>	0.06	1.87 <sup>(a),(b)</sup>	0.10						
Sonstige	T1	0.55	0.08	0.70	0.08	0.69	0.13	1.95	.164	2.43	.091 <sup>+</sup>	1.58	.209
	T2	0.61 <sup>b</sup>	0.08	0.82 <sup>a</sup>	0.08	0.46	0.14						
Freie Antworten	T1	1.92	0.13	2.28	0.13	2.16	0.22	0.55	.459	3.27	.041*	0.15	.859
	T2	1.70 <sup>b</sup>	0.13	2.14 <sup>a</sup>	0.13	2.07	0.22						

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen:

M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Gruppen

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

Auch bei den Einzelbetrachtungen bleibt das generelle Muster (mit Ausnahme der freien Antworten) bestehen, dass sich Anstiege für die Interventionsgruppen 1 und 3 und etwa gleichbleibende Ausprägungen in der Wartegruppe 2 zeigen. In insgesamt drei Bereichen finden sich signifikante Interaktionseffekte und zwar bei den Inhalten Unfallursachen, Fahrphysik und Informationsverarbeitung. Abbildung 15 zeigt den Verlauf des Wissens in diesen Bereichen für die Studiengruppen. Wissen um Unfallursachen und Informationsverarbeitung steigt durch das Training in den Gruppen 1 und 3 signifikant an, allerdings ist dieser Zuwachs gering und weist entsprechend geringe Effektstärken auf (partielles  $\eta^2=.012$  bzw.  $\eta^2=.031$ ).

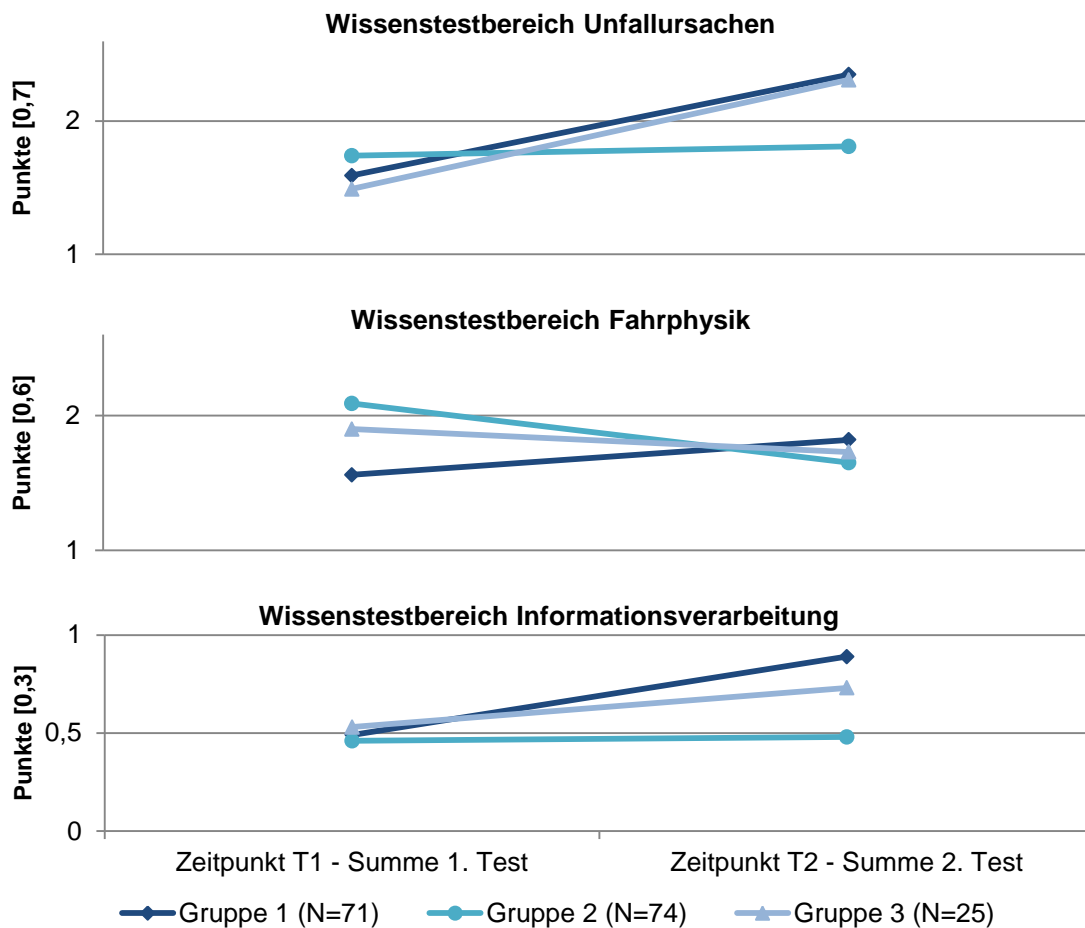


Abbildung 15: Wissenstestbereiche mit signifikanten Interaktionseffekten zwischen den Gruppen und Zeitpunkten

Im Bereich der Fahrphysik findet sich ein solcher Wissenszuwachs nur für Gruppe 1 (partielles  $\eta^2=.020$ ), während die zweite Interventionsgruppe hier einen leicht sinkenden Trend zeigt. Da alle Veränderungen auf einem geringen Niveau liegen, bleibt „Luft nach oben“. Für den Aufbau bzw. ggf. Modifikationen des Trainings sollten aber vor allem die Bereiche rechtliche Grundlagen und das Entscheidungsverhalten, in denen sich keine systematischen Unterschiede finden, betrachtet werden.

In den kontrollierten Varianzanalysen zeigte sich auch ein signifikanter Einfluss des Alters auf das Wissen ( $F_{(1,165)}=5.90$ ,  $p=.016$ ). Abbildung 16 zeigt diesen Effekt im Vergleich zwischen älteren und jüngeren Probanden (Medianschnitt von 32 Jahren) aufzeigen. Insgesamt haben die jüngeren Teilnehmer zu beiden Messzeitpunkten höhere Punktzahlen als die älteren Probanden. Der Wissenszuwachs durch das Training ist in der Interventionsgruppe 1 für jüngere und ältere Probanden zu finden, in der Gruppe 3, ohne vorherige Schichtmessungen, jedoch nur bei den jüngeren Probanden. Deskriptiv auffällig ist auch, dass die

jüngeren Teilnehmer der Wartegruppe 2 sich im Vergleich zum Vortest verschlechtern, während ihre älteren Kollegen auf dem gleichen Wissensniveau bleiben.

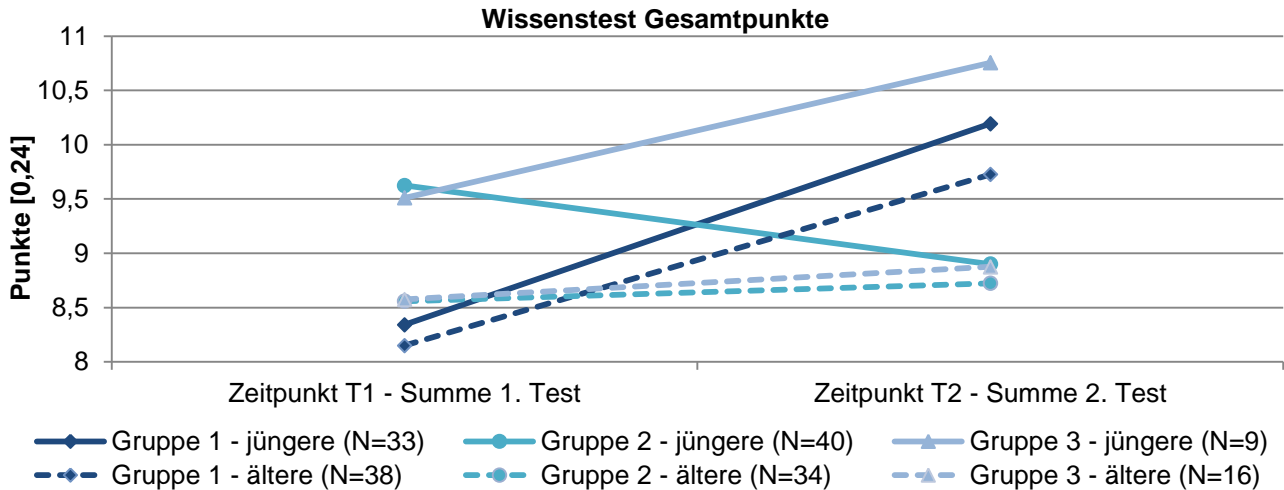


Abbildung 16: Erreichte Punktzahl im Wissenstest unterteilt nach Versuchsgruppe und Alter der Probanden

Geschlecht als Kontrollvariable hatte keinen eigenständigen Einfluss auf die Veränderungen der Gesamtpunkte im Wissenstest. Vereinzelt Unterschiede in den verschiedenen Wissensbereichen sind nicht durchgängig in eine Richtung und kaum sinnvoll interpretierbar. Da in der Gesamtanalyse (vgl. Tabelle 22) für Geschlecht kontrolliert wurde, wird von einer vertieften Analyse abgesehen.

Im Bereich *verkehrsrelevanter Persönlichkeitseigenschaften* wurden mittels adaptiertem TVP (vgl. Abschnitt 3.2.1.2) die Eigenschaften Extraversion, Sensation Seeking, Gewissenhaftigkeit, Unverträglichkeit sowie Reaktanz erfasst. Lediglich Extraversion weist einen tendenziellen Interaktionseffekt auf (vgl. Tabelle 22). Alle anderen Variablen zeigen weder eine Änderung über die Zeit noch einen Unterschied zwischen den Versuchsgruppen. Abbildung 17 zeigt den Trainingseffekt für die verkehrsbezogene Extraversion. Es zeigt sich, dass der Effekt wesentlich aus der Entwicklung der Ausprägung in der Versuchsgruppe 3 resultiert, deren Wert ansteigt, wohingegen die beiden Gruppen 1 und 2 in ihrer Ausprägung geringfügig absinken. Somit kann hier nicht von einem Trainingseffekt gesprochen werden, da sowohl Gruppe 1 als auch Gruppe 3 am Training teilgenommen haben.

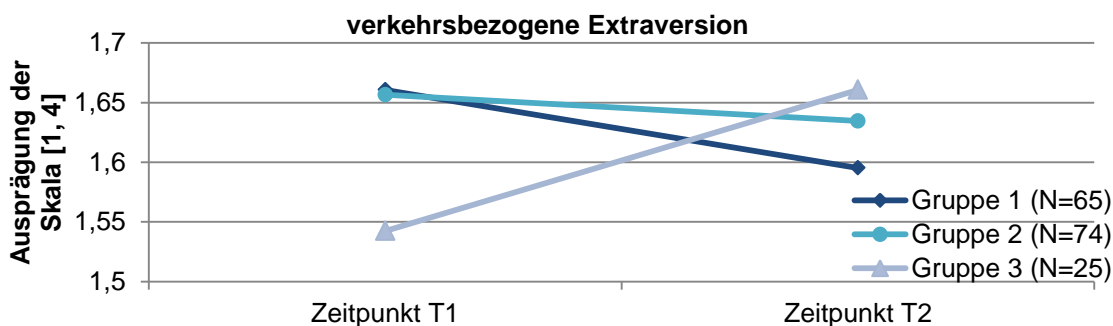


Abbildung 17: Interaktionseffekt zwischen Gruppen und Zeitpunkten bei verkehrsbezogener Extraversion

Obwohl sich damit keine Trainingseffekte im Bereich der verkehrsbezogenen Persönlichkeitseigenschaften zeigen lassen, finden sich in den Analysen nahezu durchgehend Alters- und/oder Geschlechtseffekte. Diese werden im Folgenden kurz dargestellt. Deutliche Alterseffekte weisen Extraversion ( $F_{(1,159)}=7.28$ ,  $p=.008$ ), Gewissenhaftigkeit ( $F_{(1,158)}=7.54$ ,  $p=.007$ ) und Unverträglichkeit ( $F_{(1,159)}=8.34$ ,  $p=.004$ ) auf. Tabelle 24 zeigt die Mittelwerte zu den Zeitpunkten für die drei Versuchsgruppen und einen Gesamtmittelwert. Im Gesamtmittelwert ist erkennbar, dass die jüngeren Teilnehmer sich als extrovertierter, weniger gewissenhaft und weniger verträglich beschreiben als die älteren Probanden. Dieser Unterschied ist auch in den drei Gruppen zu sehen – mit einer Ausnahme: Bei der verkehrsbezogenen Extraversion in Gruppe 3 weisen die älteren Probanden höhere Werte in auf. Abbildung 18 verdeutlicht diesen Unterschied, der jedoch nicht statistisch signifikant ist.

Tabelle 24: Alterseinfluss auf verkehrsbezogene Persönlichkeitseigenschaften

Verkehrsbezogene Persönlichkeitseigenschaften	Zeitpunkt	Gruppe 1 (N=65)		Gruppe 2 (N=74)		Gruppe 3 (N=25)		Gesamt (N=164)	
		M jung	M alt	M jung	M alt	M jung	M alt	M jung	M alt
Extraversion	T1	1.81	1.53	1.72	1.61	1.37	1.60	1.71	1.58
	T2	1.77	1.44	1.69	1.60	1.44	1.74	1.69	1.56
Gewissenhaftigkeit	T1	2.67	2.83	2.67	2.80	2.44	2.59	2.65	2.78
	T2	2.57	2.74	2.67	2.67	2.52	2.62	2.62	2.69
Unverträglichkeit	T1	2.31	2.27	2.46	2.20	2.33	2.28	2.39	2.25
	T2	2.28	2.13	2.41	2.14	2.33	2.19	2.35	2.14

Medianschnitt des Alters liegt bei 32 Jahren, Abkürzungen: M=Mittelwert

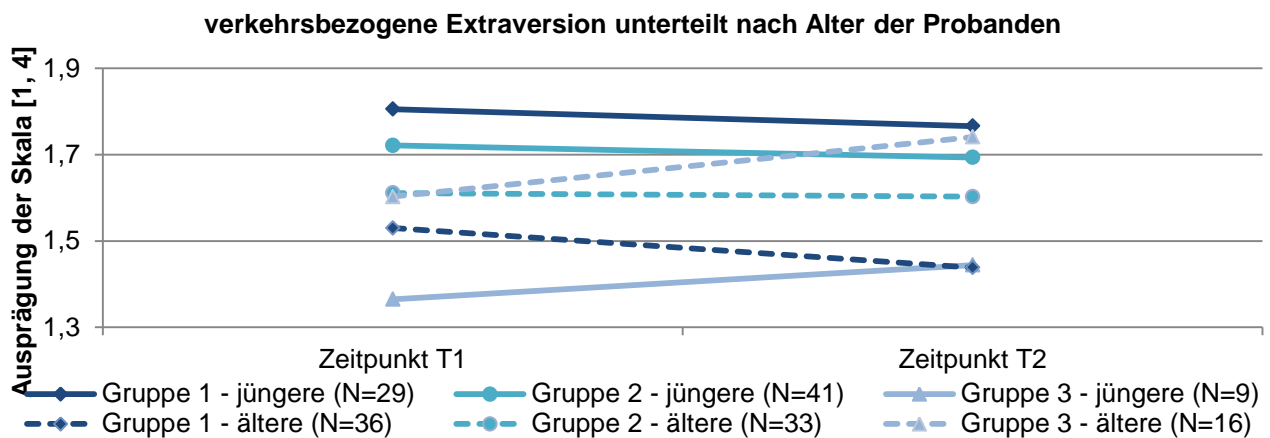


Abbildung 18: Altersabhängigkeit verkehrsbezogener Extraversion

Eine deutliche Abhängigkeit vom Geschlecht weisen die verkehrsbezogene Extraversion ( $F_{(1,159)}=8.22$ ,  $p=.005$ ), Sensation Seeking ( $F_{(1,159)}=4.25$ ,  $p=.041$ ) sowie im Trend Reaktanz ( $F_{(1,159)}=3.42$ ,  $p=.066$ ) auf. Zusätzlich weist verkehrsbezogene Gewissenhaftigkeit einen signifikanten Interaktionseffekt des Geschlechts mit den Zeitpunkten auf ( $F_{(1,158)}=4.88$ ,  $p=.029$ ). Die Mittelwerte sind in Tabelle 25 dargestellt. Die männlichen Probanden sind extrovertierter und weisen höhere Werte im Sensation Seeking auf (Ausnahme Gruppe 3). Auch die Reaktanz ist bei den Männern höher ausgeprägt (Ausnahme Gruppe 2 zum Zeitpunkt T1). Der Interaktionseffekt der Gewissenhaftigkeit wird in Abbildung 19 veranschaulicht. Er zeigt, dass Frauen hinsichtlich der Einstellung verkehrsbezogener Gewissenhaftigkeit im Gegensatz zu Männern eher von dem Training zu profitieren scheinen.

Tabelle 25: Einfluss des Geschlechts bei verkehrsbezogenen Persönlichkeitseigenschaften

Verkehrsbezogene Persönlichkeitseigenschaften	Zeitpunkt	Gruppe 1 (N=65)		Gruppe 2 (N=74)		Gruppe 3 (N=25)		Gesamt (N=164)	
		M ♀	M ♂	M ♀	M ♂	M ♀	M ♂	M ♀	M ♂
Extraversion	T1	1.56	1.68	1.46	1.73	1.50	1.52	1.50	1.68
	T2	1.45	1.62	1.51	1.69	1.63	1.63	1.51	1.65
Sensation Seeking	T1	2.08	2.36	2.24	2.43	2.39	2.36	2.21	2.39
	T2	2.17	2.33	2.26	2.39	2.27	2.31	2.23	2.35
Gewissenhaftigkeit	T1	2.87	2.73	2.63	2.76	2.36	2.61	2.66	2.73
	T2	2.90	2.61	2.69	2.66	2.62	2.57	2.75	2.63
Reaktanz	T1	1.31	1.47	1.58	1.50	1.14	1.63	1.40	1.51
	T2	1.28	1.51	1.45	1.54	1.20	1.66	1.34	1.54

Abkürzungen: M=Mittelwert

Als weiteres Maß zur Erfassung der Einstellung zur Verkehrssicherheit wurde die Risikowahrnehmung im Straßenverkehr (Delhomme, 1991; Rosenbloom et al., 2008) und das allgemeine Risk Seeking (Seipel, 2012) erfasst (vgl. Abschnitt 3.2.1.2). Die Risikowahrnehmung bei schweren Regelverstößen wurde von allen Gruppen schon zum ersten Messzeitpunkt als sehr risikoreich bewertet und somit wurde hier wie erwartet keine Veränderung über die Zeit oder durch das Training gefunden.

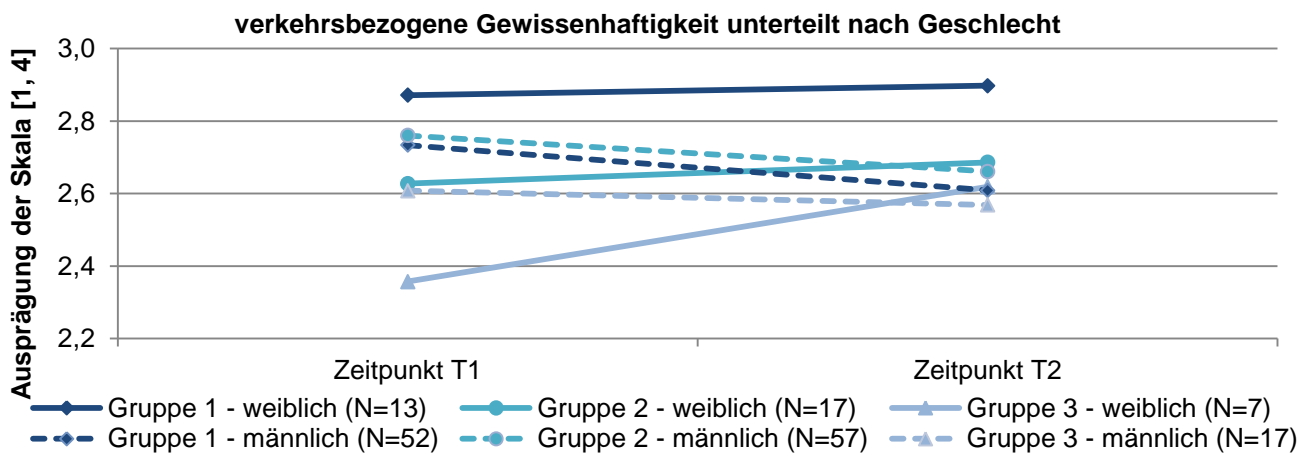


Abbildung 19: Interaktionseffekt zwischen Geschlecht und Zeitpunkten auf verkehrsbezogene Gewissenhaftigkeit

Risikowahrnehmung bei mittelschweren und leichten Regelverstößen konnte ebenfalls nicht durch das Training beeinflusst werden. Allerdings findet sich für die Risikowahrnehmung bei Normalfahrten ein signifikanter Interaktionseffekt  $p=.018$  ( $\eta^2=.049$ ) in die erwartete Richtung - die beiden Interventionsgruppen 1 und 3 bewerten normale Fahrsituationen nach dem Training als risikoreicher im Vergleich zur Wartegruppe 2 (vgl. Abbildung 20), die in ihrer Einschätzung unverändert bleibt. In den Mittelwerten ist über die verschiedenen Situationsklassen auch zu beobachten, dass die Wahrnehmung des Risikos mit abnehmenden Regelverstößen in allen Gruppen abnimmt, gleichzeitig aber nach dem Training kontinuierlich mit geringer werdendem Schweregrad zunimmt.



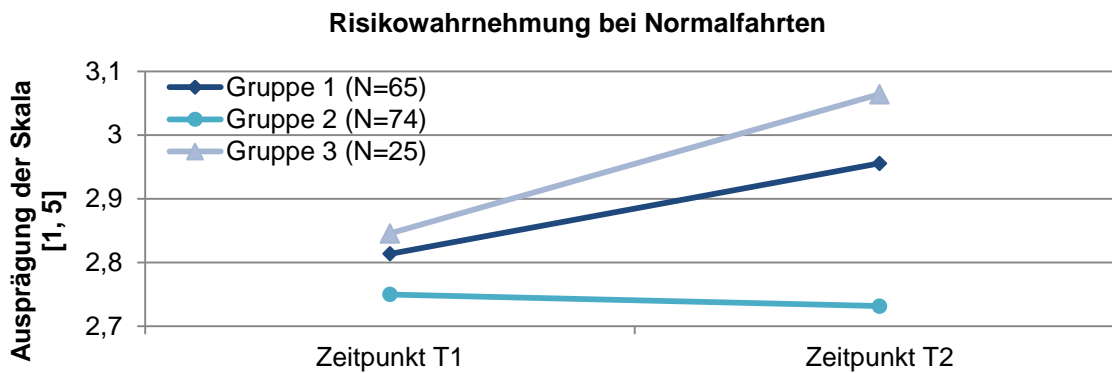


Abbildung 20: Trainingseffekt der Risikowahrnehmung bei Normalfahrten

Wie auch aus der Literatur (vgl. Abschnitt 1) und auch aus Unfallstatistiken zu erwarten weisen alle Variablen der Risikowahrnehmung und das Risk Seeking einen zum Teil hochsignifikanten Alterseinfluss auf ( $\eta^2$  zwischen .053 und .131) (vgl. Mittelwerte in Tabelle 26). Ein differenzieller Effekt des Alters auf die Trainingseffekte findet sich nicht, allerdings gibt es tendenziell eine Interaktion zwischen Alter und Zeitpunkt in der Risikowahrnehmung bei leichten Regelverstößen, in die Richtung, dass je jünger die Probanden sind, desto mehr steigt die Risikowahrnehmung zum zweiten Messzeitpunkt an.

Tabelle 26: Alterseinfluss auf Risikobewusstsein

Verkehrsbezogene Persönlichkeits- eigenschaften	Zeit- punkt	Gruppe 1 (N=65)		Gruppe 2 (N=74)		Gruppe 3 (N=25)		Gesamt (N=164)	
		M jung	M alt	M jung	M alt	M jung	M alt	M jung	M alt
Risikowahrnehmung bei schweren Regelverstößen	T1	4.47	4.43	4.37	4.50	4.53	4.53	4.43	4.48
	T2	4.43	4.58	4.40	4.49	4.40	4.61	4.41	4.55
Risikowahrnehmung bei mittelschweren Regelverstößen	T1	3.74	3.90	3.41	3.87	3.65	3.96	3.55	3.90
	T2	3.86	4.05	3.54	3.87	3.80	4.10	3.69	3.99
Risikowahrnehmung bei leichten Regelverstößen	T1	3.35	3.54	3.19	3.37	3.39	3.58	3.27	3.48
	T2	3.48	3.63	3.20	3.35	3.44	3.72	3.33	3.54
Risikowahrnehmung bei normalem Fahren	T1	2.75	2.93	2.54	2.91	2.70	2.98	2.64	2.93
	T2	2.92	3.04	2.55	2.88	2.76	3.28	2.71	3.02
Risk Seeking	T1	1.97	1.68	1.92	1.81	2.11	2.09	1.96	1.81
	T2	2.03	1.58	1.73	1.65	2.19	1.98	1.89	1.69

Medianschnitt des Alters liegt bei 32 Jahren, Abkürzungen: M=Mittelwert

Bei der allgemeinen, nicht verkehrsbezogenen Skala zum Risk Seeking findet sich zusätzlich zum Alterseffekt auch ein Haupteffekt des Geschlechts ( $F_{(1,158)}=8.93$ ,  $p=.003$ ) sowie ein tendenzieller Interaktionseffekt ( $F_{(1,158)}=2.95$ ,  $p=.088$ ). Frauen geben generell geringere Werte an als Männer und dieser Wert sinkt in allen drei Studiengruppen über die Zeit noch weiter. Bei Männern verändert sich das Risk Seeking in den Interventionsgruppen 1 und 3 nahezu nicht, während er in der Wartegruppe 2 leicht absinkt (vgl. Abbildung 21). Damit findet sich für dieses allgemeine persönlichkeitsnahen Merkmal eine nicht hypothesenkonforme Entwicklung.

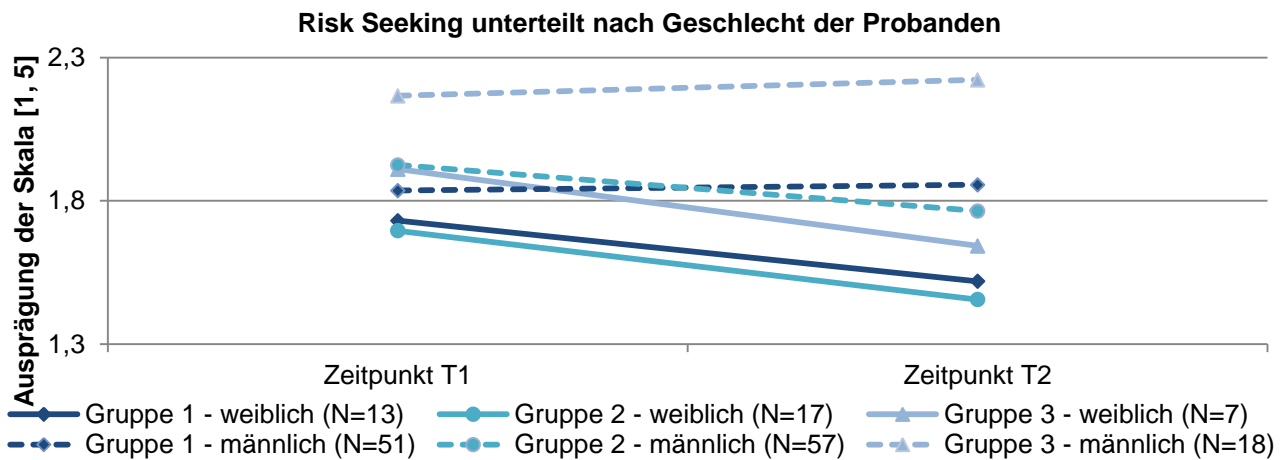


Abbildung 21: Risk Seeking im zeitlichen Verlauf in Abhängigkeit vom Geschlecht

Mit insgesamt vier Einzelskalen wurden die Einschätzung der *subjektiven Fahrfähigkeit* vorgenommen (vgl. Abschnitt 3.2.1.3). Bezogen auf die theoretischen Mittelwerte der Skalen, schätzen sich die Probanden insgesamt überdurchschnittlich gut ein bei einer etwas unterdurchschnittlichen Wahrscheinlichkeit, in einen Unfall verwickelt zu werden. Die Differenzierung zwischen allgemeinem Straßenverkehr und Fahrten im Rettungsdienst macht dabei kaum einen Unterschied (vgl. Abbildung 22).

Im Hinblick auf die Studiengruppen sind die Verläufe der subjektiven Fahrfähigkeit nicht eindeutig zu interpretieren. Signifikante Interaktionseffekte gibt es in diesem Evaluationsbereich nur für die allgemeine Fahrkompetenz - allerdings schätzt sich nur Gruppe 1 nach dem Training als kompetenter ein, wohingegen die anderen beiden Gruppen gleichbleiben bzw. minimal absinken. Hinsichtlich der Unfallwahrscheinlichkeit scheinen sich die beiden Gruppen, die das Training bekommen haben, deskriptiv sogar genau entgegengesetzt zu entwickeln. Da die Effekte jedoch nicht signifikant sind und die Unterschiede sehr klein, können diese Entwicklungen auch auf Zufall zurückgehen. Generell ist ein positiver Effekt des Trainings in diesem Bereich also nicht auszumachen.

Allerdings finden sich auch bei der Einschätzung der eigenen Fahrfähigkeit Unterschiede hinsichtlich Geschlecht und Alter. Für die wahrgenommene allgemeine Unfallwahrscheinlichkeit gibt es einen signifikanten Interaktionseffekt ( $F_{(1,158)}=3.91, p=.050$ ) mit dem Geschlecht und für die wahrgenommene Unfallwahrscheinlichkeit im Rettungsdienst einen tendenziellen Interaktionseffekt mit dem Alter ( $F_{(1,158)}=3.86, p=.057$ ) (vgl. Abbildung 23).

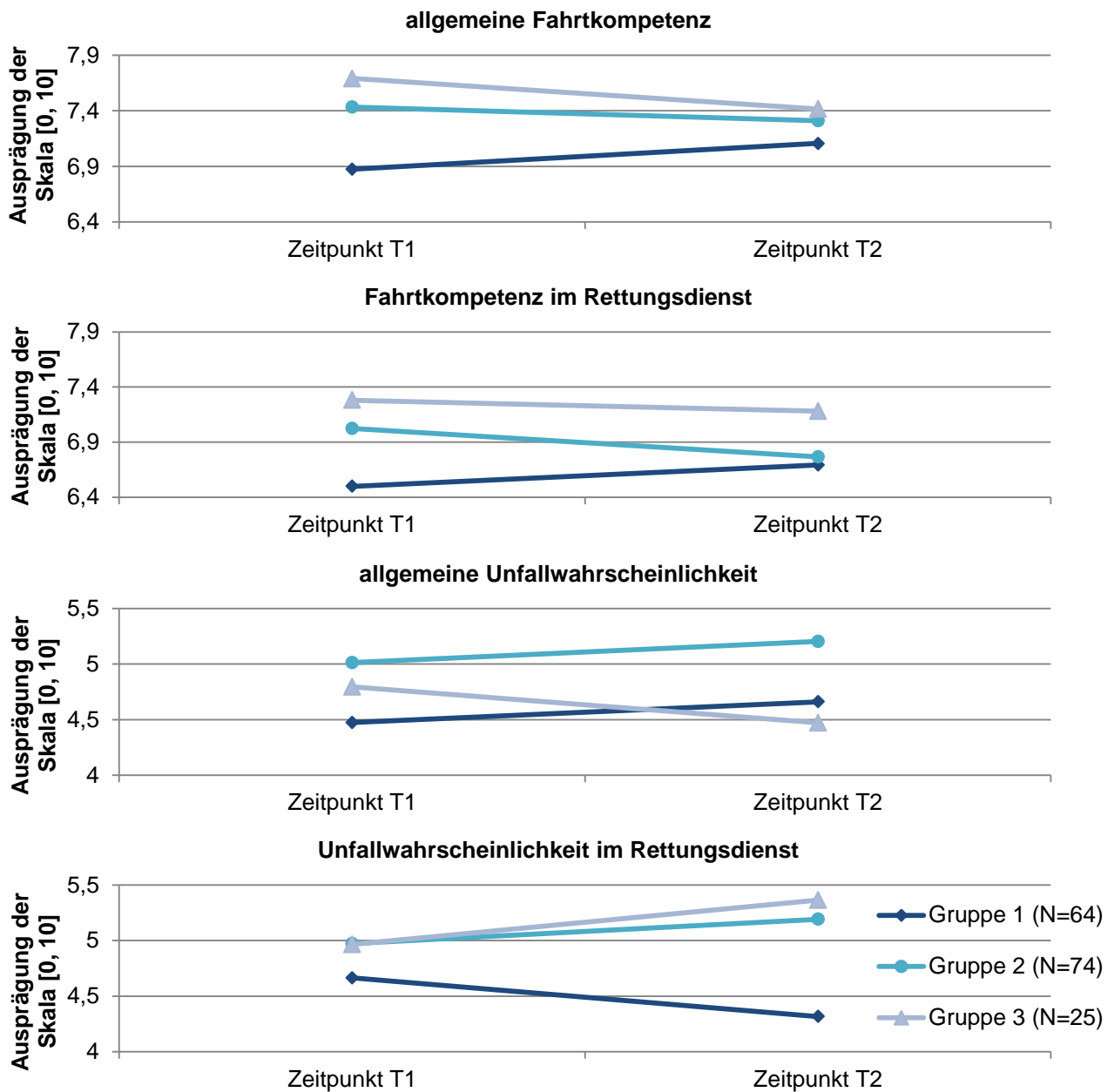


Abbildung 22: Fahrfähigkeit (alle Skalen) im zeitlichen Verlauf

Frauen scheinen unabhängig von der Studiengruppe - also auch unabhängig von der Trainingsteilnahme - ihre eigene allgemeine Unfallwahrscheinlichkeit bei der zweiten Messung höher zu bewerten, Männer bewerten sie gleich (Gruppen 1 und 2) oder sogar geringer (Gruppe 3). Die subjektive Unfallwahrscheinlichkeit im Rettungsdienst entwickelt sich hingegen sehr unterschiedlich. Die älteren Probanden der Interventionsgruppe 1 und der Wartegruppe 2 schätzen ihre Unfallwahrscheinlichkeit zum Zeitpunkt T2 geringer ein als vorher, alle anderen Gruppen bleiben gleich oder schätzen sie etwas höher ein. Wie schon in den kontrollierten varianzanalytischen Ergebnissen (vgl. Tabelle 22) zu sehen, sind für diesen Variablenbereich also keine Trainingseffekte zu finden.

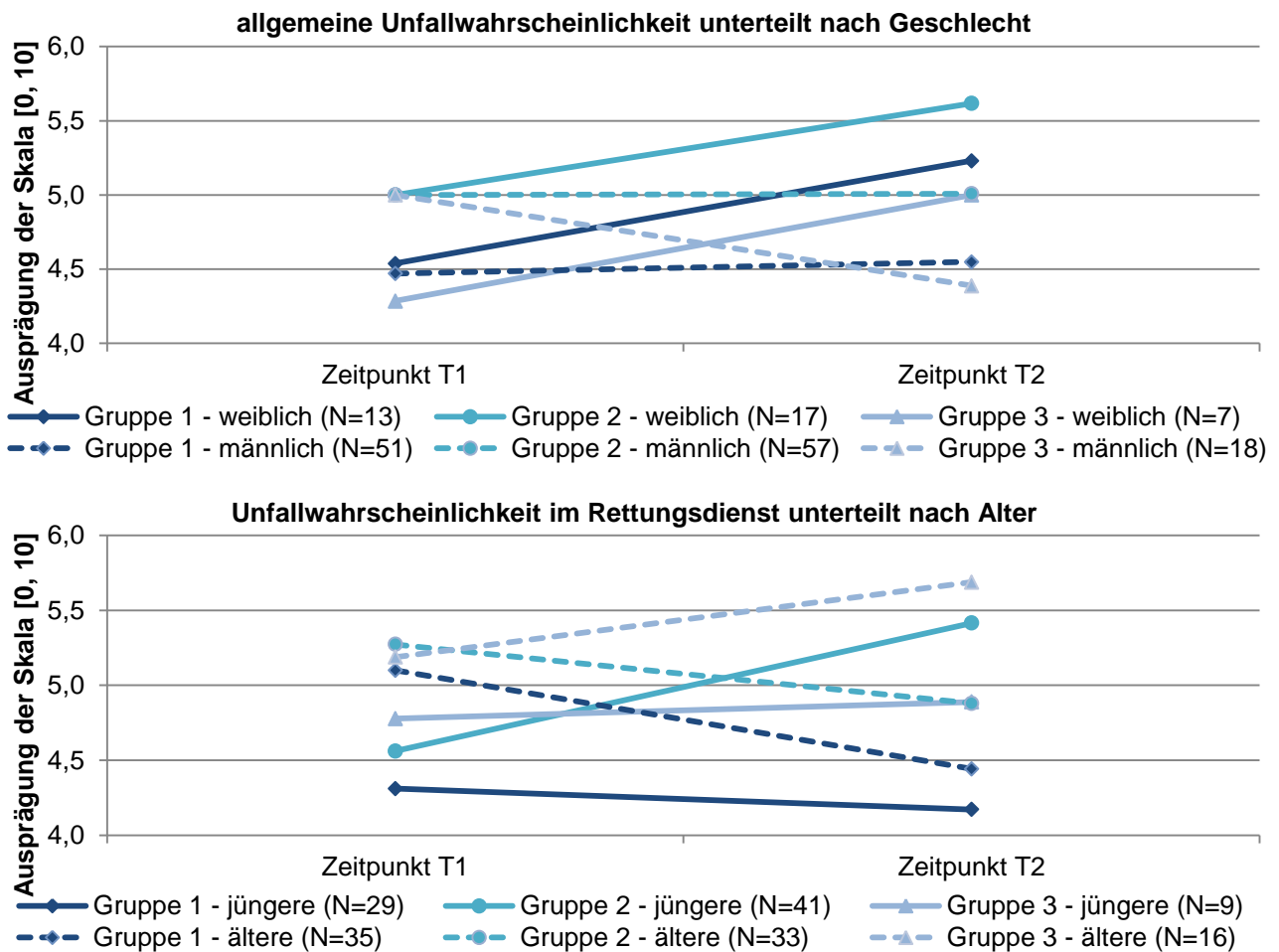


Abbildung 23: Verlauf der wahrgenommenen eigenen Unfallwahrscheinlichkeit in Abhängigkeit von Alter und Geschlecht

#### 4.2.2.2 Haupteffekte in der Verhaltensebene

Im Folgenden werden die Hauptanalysen für die Evaluationsebene des Verhaltens dargestellt. Da in Gruppe 3 diese Messungen zum Zeitpunkt T1 nicht durchgeführt wurden, können alle Berechnungen der Zeit- und Testeffekte nur für die Gruppen 1 und 2 durchgeführt werden. Haupt- und Interaktionseffekte finden sich in Tabelle 27. Für die Fahrprofilaten werden hierbei lediglich die RTW-Fahrten zum Einsatzort genutzt (daher zum Teil auch kleineres N), um die Vergleichbarkeit innerhalb der aggregierten Daten zu erhalten und um nur diejenigen Fahrten zu nutzen, die primärer Zielbereich des Trainings sind. Alle Effekte werden im Weiteren genauer beschrieben.

In die Analysen konnten 59 Probanden der Gruppe 1 und 60 Probanden der Gruppe 2 mit RTW-Fahrten zum Einsatzort zu beiden Zeitpunkten aufgenommen werden. Die durchschnittliche Dauer der in die Analysen eingeschlossenen Fahrten war zu T1 in dieser Gruppe 66:14 Minuten (Min=586 Sek., Max=9644 Sek., SD=1873 Sek.) und zu T2 53:30 Minuten (Min=759 Sek., Max=7968 Sek., SD=1690 Sek.). Für Gruppe 2 war die durchschnittliche Dauer der eingeschlossenen RTW-Fahrten zu T1 68:20 Minuten (Min=516 Sek., Max=10294 Sek., SD=2384 Sek.) und für T2 53:53 Minuten (Min=226 Sek., Max=8284 Sek., SD=2000 Sek.). Die berechneten Anzahlen und Dauer der Trigger für Längs- und Querschleunigung wurden jeweils um die Dauer der Einsatzfahrten bereinigt. Alle Fahrten sind eine Mischung aus inner- und außerörtlichen Fahrten, die sich zwischen den Gruppen nicht systematisch unterscheiden und zwischen den Messzeitpunkten konstant bleiben, da nur sehr wenige Probanden zwischen den Zeitpunkten ihre Wache gewechselt haben und somit die Örtlichkeiten im Längsschnitt größtenteils identisch sind.

Tabelle 27: Haupteffekte der Evaluationsebene „Verhalten“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
			M	SE	M	SE	F	p	F	p	F	p
Fahrprofilaten: Geschwindigkeiten												
Durchschnittliche Geschwindigkeit (in km/h)	119	T1	54.98	1.38	51.47	1.37	0.80	.372	1.38	.242	3.75	.055 <sup>+</sup>
		T2	53.42	1.46	52.44	1.45						
Maximale Geschwindigkeit (in km/h)	119	T1	117.79	2.58	115.01	2.56	1.14	.289	0.00	.997	3.62	.059 <sup>+</sup>
		T2	112.72	2.81	115.46	2.79						
Fahrprofilaten: Längsbeschleunigung												
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	119	T1	3.65	0.27	3.38	0.27	1.55	.216	0.00	.965	1.26	.264
		T2	2.62	0.14	2.88	0.14						
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	119	T1	0.31	0.02	0.33	0.02	1.75	.189	0.49	.484	0.67	.416
		T2	0.27	0.01	0.27	0.01						
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	119	T1	0.0157	0.02	0.0510	0.02	1.42	.236	1.22	.272	1.24	.268
		T2	0.0006	0.00	0.0005	0.00						
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	119	T1	0.0055	0.01	0.0280	0.01	1.04	.310	1.69	.197	1.64	.203
		T2	0.0003	0.00	0.0005	0.00						
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	119	T1	-3.34	0.21	-3.70	0.21	5.14	.025*	0.95	.331	1.07	.304
		T2	-3.08	0.14	-3.05	0.14						
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	119	T1	-0.32	0.02	-0.34	0.02	0.63	.428	0.16	.692	1.18	.279
		T2	-0.29	0.01	-0.28	0.01						
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	119	T1	0.763	0.48	1.420	0.48	1.39	.240	0.73	.396	1.13	.290
		T2	0.139	0.04	0.064	0.04						
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	119	T1	0.235	0.19	0.524	0.19	1.23	.269	1.11	.295	1.17	.282
		T2	0.016	0.00	0.011	0.00						
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	119	T1	0.005	0.02	0.053	0.02	1.29	.259	2.06	.154	2.70	.103
		T2	0.004	0.00	0.001	0.00						
Anzahl Trigger pro Minute negative Längsbeschleunigung (5,9 m/s <sup>2</sup> )	119	T1	0.003	0.01	0.028	0.01	1.14	.287	2.20	.141	2.47	.119
		T2	0.001	0.00	0.000	0.00						
Fahrprofilaten: Querschleunigung												
Maximale positive Querschleunigung (in m/s <sup>2</sup> )	119	T1	5.38	0.47	5.19	0.47	1.11	.293	0.23	.634	0.00	.960
		T2	3.99	0.15	3.83	0.15						
Mittlere positive Querschleunigung (in m/s <sup>2</sup> )	119	T1	0.51	0.04	0.49	0.04	0.01	.931	0.29	.591	0.00	.995
		T2	0.41	0.02	0.39	0.02						
Maximale negative Querschleunigung (in m/s <sup>2</sup> )	119	T1	-4.71	0.44	-4.96	0.43	1.18	.279	0.02	.884	0.39	.535
		T2	-3.92	0.16	-3.77	0.16						
Mittlere negative Querschleunigung (in m/s <sup>2</sup> )	119	T1	-0.47	0.04	-0.46	0.04	0.18	.674	0.09	.771	0.03	.873
		T2	-0.39	0.02	-0.38	0.02						
Dauer Trigger Querschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	119	T1	0.432	0.44	1.342	0.44	1.93	.167	2.13	.147	2.10	.150
		T2	0.002	0.00	0.005	0.00						

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
			M	SE	M	SE	F	p	F	p	F	p
Anzahl Trigger pro Minute Querschleunigung (6,9 m/s <sup>2</sup> )	119	T1	0.211	0.18	0.563	0.18	2.08	.152	1.82	.180	1.79	.184
		T2	0.000	0.00	0.001	0.00						
Selbsteinschätzung der Einsatzfahrten												
Fahrerbeurteilung sichere Fahrweise	145	T1	82.02	1.33	82.12	1.29	0.16	.692	0.11	.736	0.44	.508
		T2	79.54	1.77	78.17	1.70						
Fahrerbeurteilung vorausschauende Fahrweise	144	T1	80.64	1.29	82.75	1.26	0.24	.628	0.31	.577	1.69	.196
		T2	80.86	1.56	80.79	1.52						
Fahrerbeurteilung rasante Fahrweise	144	T1	42.69	2.46	36.37	2.39	0.07	.799	2.30	.131	0.91	.341
		T2	41.75	2.60	38.29	2.53						
Fahrerbeurteilung hektische Fahrweise	144	T1	19.44	1.67	13.89	1.63	0.25	.617	4.51	.035*	0.79	.376
		T2	20.55	1.76	16.71	1.71						
Fremdbeurteilung der Einsatzfahrten												
Beifahrerbeurteilung sichere Fahrweise	120	T1	87.73	1.63	87.72	1.69	0.90	.346	0.31	.576	0.41	.525
		T2	88.65	1.43	90.53	1.48						
Beifahrerbeurteilung vorausschauende Fahrweise	120	T1	87.04	1.43	85.96	1.48	2.30	.132	0.03	.853	0.32	.570
		T2	87.69	1.47	88.17	1.52						
Beifahrerbeurteilung rasante Fahrweise	120	T1	35.11	2.65	34.84	2.74	0.20	.658	0.18	.675	0.21	.651
		T2	37.41	2.89	34.96	3.00						
Beifahrerbeurteilung hektische Fahrweise	120	T1	15.81	1.96	14.41	2.03	0.00	.967	0.94	.333	0.24	.624
		T2	17.43	2.28	14.13	2.36						

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen:

N=Anzahl Probanden, M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau:

+p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Fahrprofildaten enthalten lediglich RTW-Fahrten zum Einsatzort

Über alle erfassten Variablen hinweg, gibt es auf der Verhaltensebene nur für die *Geschwindigkeit* signifikante Interaktionseffekte. Hier wurden die Trainingseffekte auf die durchschnittliche und maximale Geschwindigkeit während der RTW-Fahrten berechnet. Es gibt keine Unterschiede zwischen den Gruppen oder über die Zeit hinweg, jedoch mit 10% Irrtumswahrscheinlichkeit signifikante Interaktionseffekte ( $F_{(1,115)mittel}=3.75, p=.055$ ;  $F_{(1,115)max}=3.63, p=.059$ ) mit eher kleinen Effektstärken (beide Geschwindigkeitsmaße  $\eta_{part}^2=.03$ ), die in der Abbildung graphisch dargestellt werden.

Deutlich zu erkennen ist für die Interventionsgruppe 1 eine Abnahme sowohl der mittleren als auch der maximalen Geschwindigkeit, die auf Hinfahrten zum Einsatzort erreicht wurden, im Vergleich zur Wartegruppe 2, deren Geschwindigkeiten leicht steigen oder gleich bleiben. Das Training wirkt sich demzufolge positiv auf die Geschwindigkeit aus.

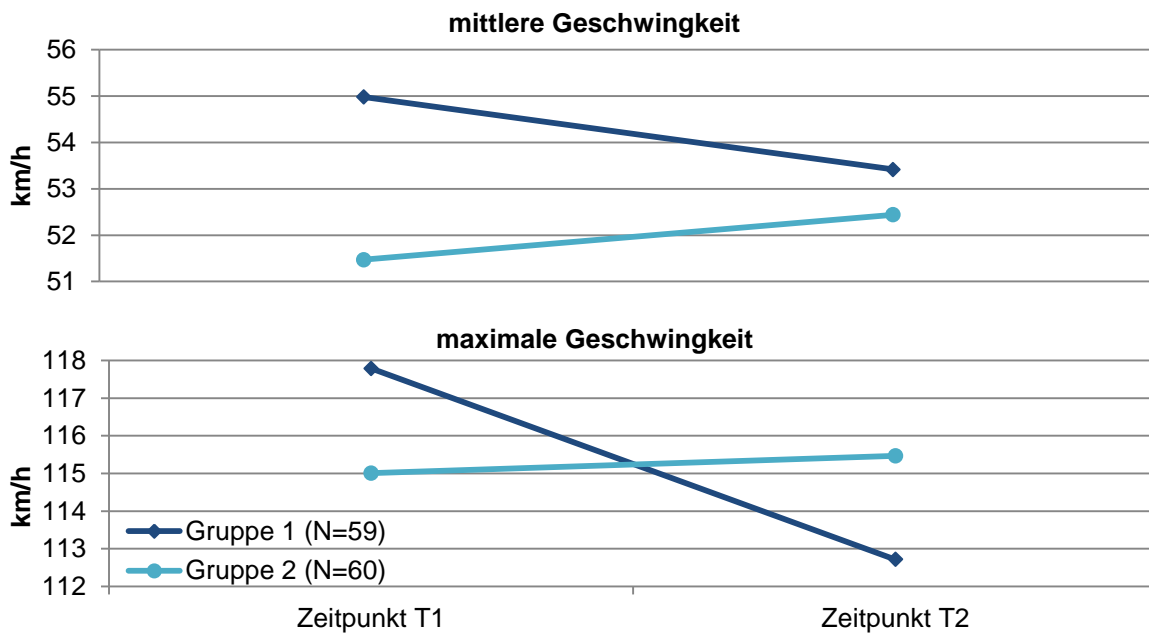


Abbildung 24: Mittlere und maximale Geschwindigkeit der RTW-Fahrten zum Einsatzort kontrolliert für Alter und Geschlecht

Ähnlich wie bei den Analysen der Lernebene, finden sich auch hier Effekte der kontrollierten Variablen: Bei der maximalen Geschwindigkeit gibt es einen schwachen Interaktionseffekt des Geschlechts mit den Zeitpunkten ( $F_{(1,115)}=3.24, p=.075$ ). Dieser Effekt geht in die Richtung, dass Männer, die das Training absolviert haben, eher ihre maximale Geschwindigkeit reduzieren als ihre weiblichen Kolleginnen (vgl. Abbildung 25). Effekte des Alters treten nicht auf.

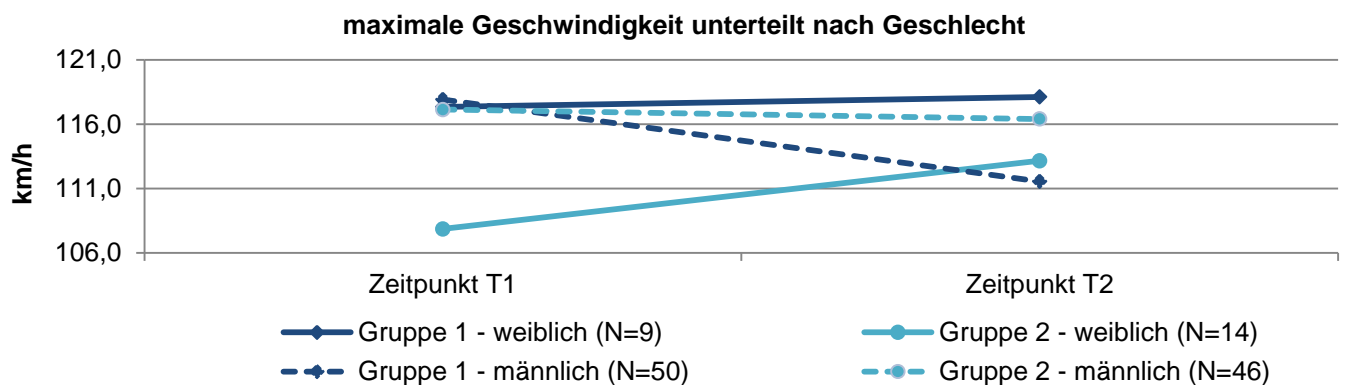


Abbildung 25: Maximale Geschwindigkeit bei RTW-Fahrten zum Einsatzort in Abhängigkeit vom Geschlecht

In den verschiedenen untersuchten Aspekten der *Längsbeschleunigung* finden sich keine Interaktionseffekte, die auf positive Trainingseffekte hindeuten. Allerdings finden sich signifikante Zeiteffekte für alle abhängigen Variablen der Längsbeschleunigung, wenn keine Kontrolle für Alter und Geschlecht vorgenommen wird, dahingehend, dass sowohl in der Interventions- wie in der Wartegruppen die (positiven und negativen) Beschleunigungen zu T2 geringer sind als zu T1. Dies scheint im Wesentlichen durch das Geschlecht beeinflusst zu sein, und wird im Folgenden noch einmal genauer dargestellt. Tabelle 28 zeigt die Mittelwerte der verschiedenen Längsbeschleunigungen sowie den Haupt- und Interaktionseffekt des Geschlechts.

Tabelle 28: Einfluss des Geschlechts auf die Längsbeschleunigungen

Abhängige Variable	Zeitpunkt	Gruppe 1 (N=59)		Gruppe 2 (N=60)		Geschlechtseffekt		Interaktionseffekt	
		M ♀	M ♂	M ♀	M ♂	F <sub>(1,115)</sub>	p	F <sub>(1,115)</sub>	p
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	T1	2.84	3.83	2.61	3.56	3.08	.082 <sup>+</sup>	3.10	.081 <sup>+</sup>
	T2	2.29	2.65	3.28	2.79				
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	T1	0.24	0.33	0.28	0.34	4.37	.039*	4.01	.048*
	T2	0.26	0.28	0.28	0.27				
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	T1	0.00	0.02	0.00	0.06	0.75	.389	0.84	.361
	T2	0.00	0.00	0.00	0.00				
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	T1	0.00	0.01	0.00	0.03	0.71	.402	0.90	.344
	T2	0.00	0.00	0.00	0.00				
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	T1	-2.65	-3.50	-2.73	-3.96	3.21	.076 <sup>+</sup>	9.41	.003*
	T2	-3.17	-3.02	-3.57	-2.94				
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	T1	-0.25	-0.33	-0.29	-0.35	3.81	.053 <sup>+</sup>	4.24	.042*
	T2	-0.27	-0.29	-0.29	-0.28				
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	T1	0.06	1.00	0.10	1.70	1.88	.173	1.82	.180
	T2	0.08	0.15	0.11	0.05				
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	T1	0.01	0.31	0.01	0.64	1.56	.214	1.62	.206
	T2	0.01	0.02	0.02	0.01				
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	T1	0.00	0.01	0.00	0.06	0.67	.415	0.93	.336
	T2	0.01	0.00	0.00	0.00				
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	T1	0.00	0.01	0.00	0.03	0.85	.358	0.81	.369
	T2	0.00	0.00	0.00	0.00				

Berechnung: Varianzanalyse mit Messwiederholung; Abkürzungen: M=Mittelwert, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001;

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Keines der Triggermaße wird durch das Geschlecht der Teilnehmer signifikant beeinflusst, dafür jedoch alle Maße der positiven und negativen Längsbeschleunigung. Abbildung 26 zeigt diese Effekte. Es ist insgesamt zu erkennen, dass die männlichen Probanden zum Zeitpunkt T1 höhere Längsbeschleunigungen aufweisen. Sie geben also stärker Gas und bremsen stärker als ihre weiblichen Kollegen. In den Mittelwerten verschwindet dieser Effekt jedoch zum Zeitpunkt T2, wo sich die beiden Geschlechtergruppen einander annähern. Bei den Maximalwerten ist der Effekt etwas anders. Bei der positiven Längsbeschleunigung steigt der Wert zum Zeitpunkt T2 bei den Frauen der Wartegruppe an, bei allen anderen Gruppen sinkt er. Bei der maximalen negativen Längsbeschleunigung (Bremsen) weisen die Frauen beider Versuchsgruppen zum Zeitpunkt T2 höhere Werte als die Männer auf. Das Verhältnis der maximalen Bremsstärken dreht sich um. Wie schon aus den fehlenden Interaktionseffekten der Hauptanalysen (vgl. Tabelle 27) zu erkennen, gibt es hier keine Trainingseffekte – beide Gruppen entwickeln sich mit Ausnahme der maximalen positiven Längsbeschleunigung jeweils gleichsinnig.



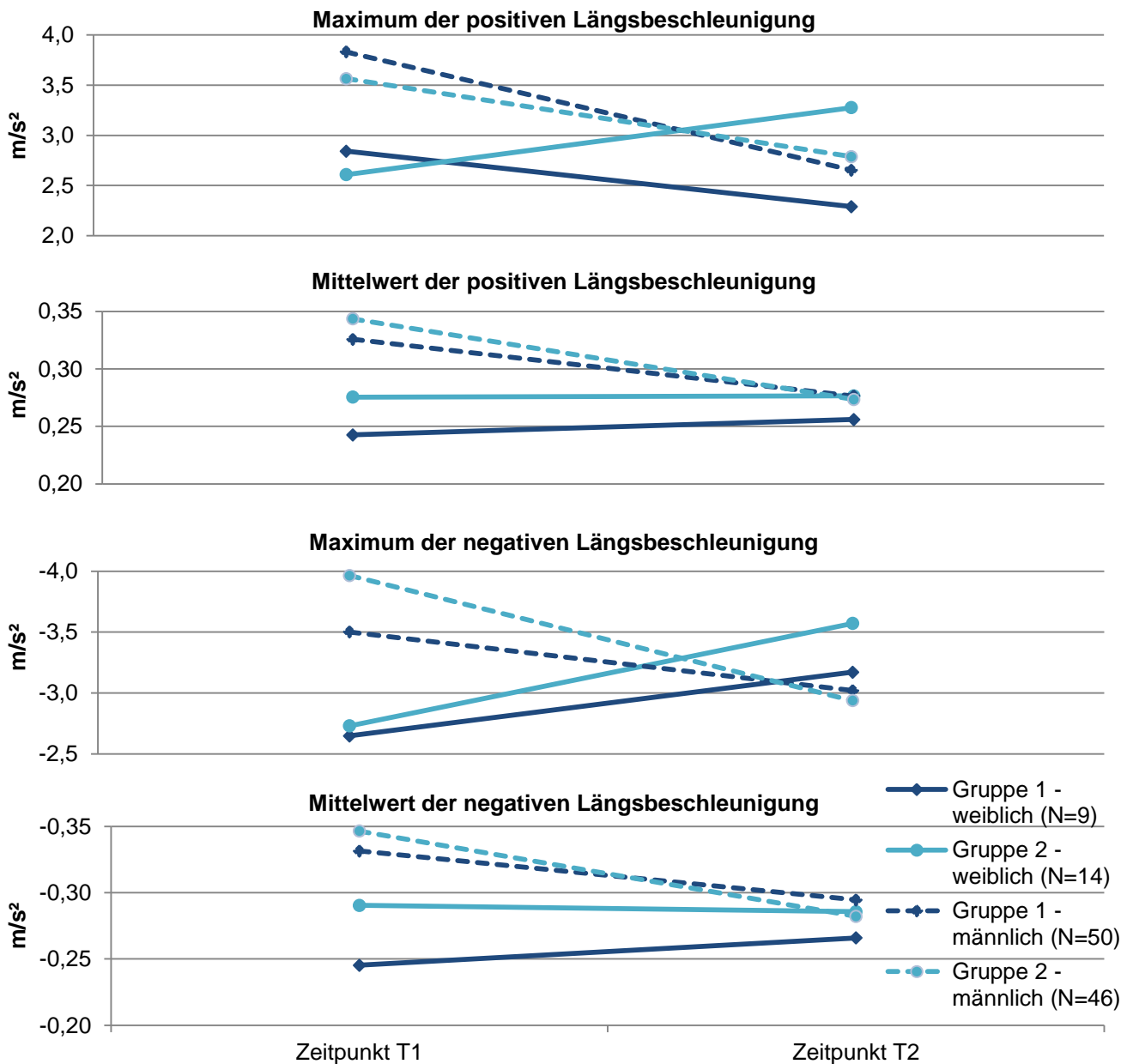


Abbildung 26: Längsbeschleunigungen bei RTW-Fahrten zum Einsatzort in Abhängigkeit von Geschlecht und Gruppe

Die in den Fahrprofilen ebenfalls betrachteten *Querbeschleunigungen* (Kurvenfahrten) verhalten sich ähnlich der Längsbeschleunigungen. Es gibt keinerlei signifikante Effekte, weder über die Zeit, noch zwischen den Gruppen oder in der Interaktion dieser beiden. Es kann also nicht von einem Trainingseffekt in diesem Bereich ausgegangen werden. Ohne Kontrolle von Alter und Geschlecht finden sich jedoch auch für die Querbeschleunigungen signifikante Zeiteffekte dahingehend, dass die Beschleunigungen über die Zeit in beiden Gruppen geringer werden.

Die Maximalwerte der Querbeschleunigungen weisen wieder signifikante Interaktionseffekte zwischen Geschlecht und der Zeit auf. Tendenzielle Haupteffekte des Geschlechts sind für die positiven Querbeschleunigungen und die maximale negative Querbeschleunigung zu finden. Für die Triggervariablen zeigen sich erneut keinerlei Effekte und sie liegen insgesamt auch einem sehr niedrigen Niveau (vgl. Tabelle 29).

Abbildung 27 zeigt, dass die männlichen Probanden zum Zeitpunkt T1 insgesamt höhere Querbeschleunigungen aufweisen als die weiblichen Probanden. Dieser Effekt wird jedoch über die Zeit hinweg aufgehoben, denn zu T2 nähern sich die Männer und Frauen wieder einander an. Auch die Art des Verlaufs ist

wie schon bei der Längsbeschleunigung für die männlichen Probanden beider Versuchsgruppen völlig identisch. Bei den Frauen sinkt die Querbewegung (bis auf die maximale positive) bei den Frauen der Wartegruppe, die der Trainingsgruppe bleibt über die Zeit hinweg stabil auf dem gleichen Niveau. Auch bei der Betrachtung der Geschlechtergruppen ist daher bei der Querbewegung kein Trainingseffekt zu finden.

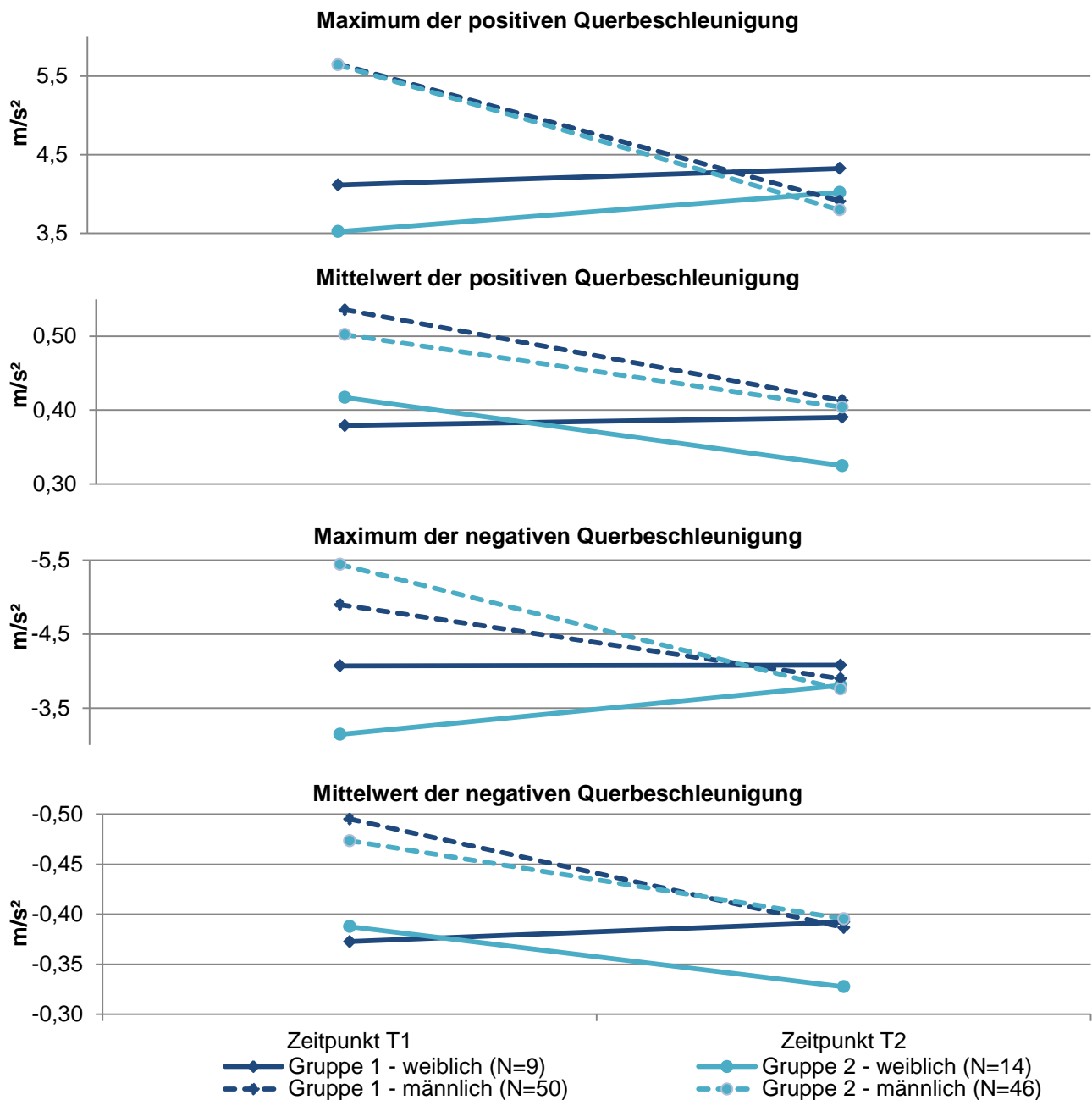


Abbildung 27: Querbewegungen bei RTW-Fahrten zum Einsatzort in Abhängigkeit von Geschlecht und Gruppe

Zusätzlich zu objektiven Messdaten während der Einsatzfahrten sollten die Probanden selbst ihre eigene Fahrleistung über die Schicht hinweg beurteilen. Dafür standen ihnen vier Skalen zu Verfügung. Die gleichen Skalen sollten auch von den Beifahrern beurteilt werden (vgl. Abschnitt 3.2.1.3; *Selbst- und Fremdeinschätzung der Einsatzfahrten*). Die Fahrweise sollte hinsichtlich sicherer, vorausschauender, rasanter und hektischer Fahrweise beurteilt werden. Insgesamt kann, bis auf eine Ausnahme, über alle acht Variablen weder ein Unterschied zwischen den Gruppen, noch über die Zeit hinweg oder eine Interaktion der Zeit und Gruppe gefunden werden. Lediglich die eigene Einschätzung der Fahrweise hinsichtlich Hektik

weist einen Unterschied zwischen den Gruppen auf (Gruppe 1 bewertet ihre eigene Fahrweise hektischer als Gruppe 2). Es gibt insgesamt jedoch keinen Trainingseffekt.

Tabelle 29: Einfluss des Geschlechts auf die Querbeschleunigungen

Abhängige Variable	Zeitpunkt	Gruppe 1 (N=59)		Gruppe 2 (N=60)		Geschlechtseffekt		Interaktionseffekt	
		M ♀	M ♂	M ♀	M ♂	F <sub>(1,115)</sub>	p	F <sub>(1,115)</sub>	p
Maximale positive Querbeschleunigung (in m/s <sup>2</sup> )	T1	4.11	5.66	3.52	5.64	3.47	.065 <sup>+</sup>	6.94	.010 <sup>**</sup>
	T2	4.32	3.91	4.02	3.80				
Maximale positive Querbeschleunigung (in m/s <sup>2</sup> )	T1	0.38	0.54	0.42	0.50	3.37	.069 <sup>+</sup>	0.91	.341
	T2	0.39	0.41	0.32	0.40				
Maximale negative Querbeschleunigung (in m/s <sup>2</sup> )	T1	-4.07	-4.90	-3.14	-5.44	3.44	.066 <sup>+</sup>	4.94	.028 <sup>*</sup>
	T2	-4.08	-3.90	-3.81	-3.76				
Maximale negative Querbeschleunigung (in m/s <sup>2</sup> )	T1	-0.37	-0.50	-0.39	-0.47	2.53	.115	0.95	.332
	T2	-0.39	-0.39	-0.33	-0.40				
Dauer Trigger Querbeschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	T1	0.03	0.63	0.00	1.62	1.55	.216	1.56	.214
	T2	0.01	0.00	0.00	0.01				
Dauer Trigger Querbeschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	T1	0.01	0.30	0.00	0.68	1.61	.208	1.62	.205
	T2	0.00	0.00	0.00	0.00				

Berechnung: Varianzanalyse mit Messwiederholung; Abkürzungen: M=Mittelwert, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001;

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

Unabhängig vom Evaluationsdesign ist auffällig (vgl. Tabelle 27), dass die Beifahrerbeurteilungen entgegen der Vermutung besser ausfallen als die Selbstbeurteilungen der Fahrer. Dieser Effekt ist besonders bei der sicheren und vorausschauenden Fahrweise deutlich zu sehen. Bei der rasanten und hektischen Fahrweise ist dieser Unterschied etwas geringer ausgeprägt.

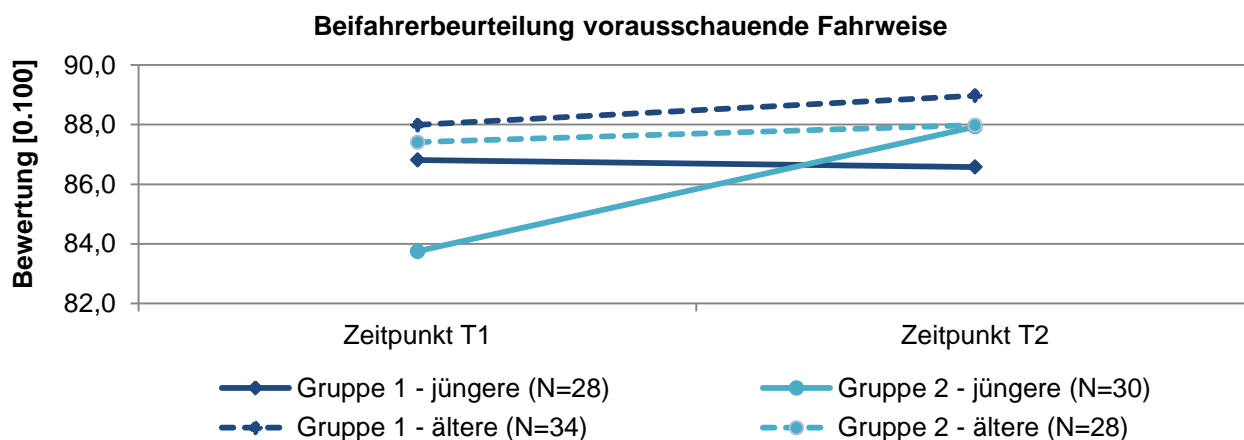


Abbildung 28: Beifahrerbeurteilung vorausschauender Fahrweise in Abhängigkeit vom Alter des Fahrers

Innerhalb des Evaluationsdesigns gibt es als weitere Einflussfaktoren lediglich einen Gesamtalterseinfluss auf die Beifahrerbeurteilung zur vorausschauenden Fahrweise (F<sub>(1,116)</sub>=5.79, p=.018) - die Beifahrer bewerten die älteren Probanden insgesamt als vorausschauender als ihre jüngeren Kollegen (vgl. Abbildung 28) -

und einen Interaktionseffekt mit dem Alter bei der Beifahrerbeurteilung zur rasanten Fahrweise ( $F_{(1,116)}=4.80$ ,  $p=.030$ ) - ältere Fahrer werden zum Zeitpunkt T2 besser bewertet als zu T1, wo ältere und jüngere gleich bewertet werden (vgl. Abbildung 29). Dieser Effekt verläuft parallel für Interventions- und Wartegruppe, so dass sich auch hier kein differenzieller Effekt des Trainings zeigt.

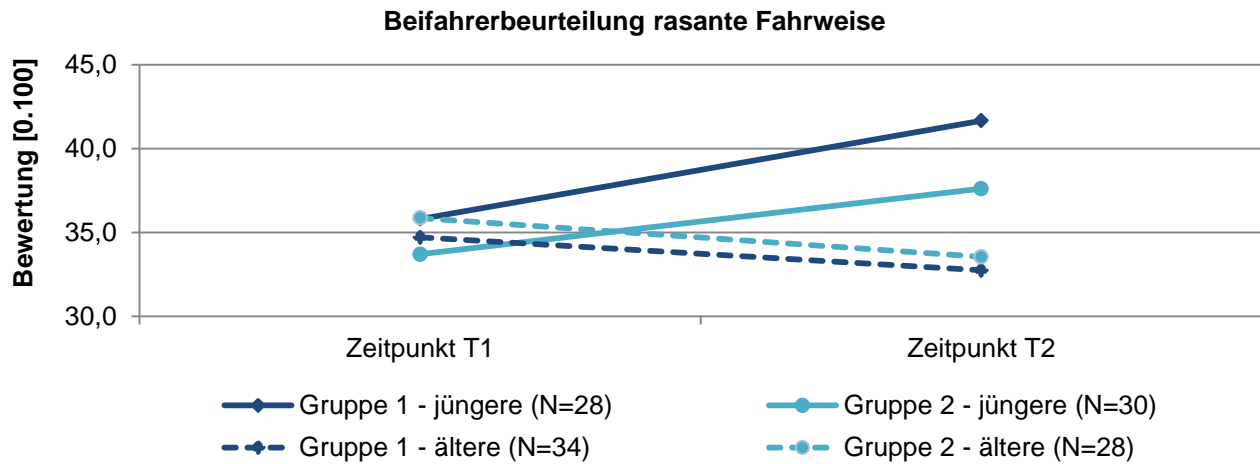


Abbildung 29: Beifahrerbeurteilung rasanter Fahrweise in Abhängigkeit vom Alter des Fahrers

#### 4.2.2.3 Haupteffekte in der Resultateebene

Auf der vierten Evaluationsebene, der Resultateebene, werden zum einen längerfristige Beanspruchungsfolgen untersucht, die für alle drei Gruppen vorliegen, und zum anderen schichtbezogene Beanspruchungen, die zum ersten und zweiten Messzeitpunkt nur vollständig für die Gruppen 1 und 2 vorliegen. Als weiteres potenzielles Resultat des Trainings werden als Proxy für die Hilfsfrist, die Länge der Einsatzfahrten betrachtet. Bei diesem Aspekt ist zu berücksichtigen, dass eine „Nichtveränderung“ als wünschenswertes Ergebnis gesehen wird. Tabelle 30 beinhaltet alle Haupt- und Interaktionseffekte sowie die um die Alters- und Geschlechtseffekte bereinigten Mittelwerte der (drei) Versuchsgruppen.

Über alle erfassten Bereiche dieser Evaluationsebene hinweg zeigen sich keinerlei Interaktionseffekte, die auf einen Trainingseffekt hindeuten würden. Allerdings ist auch zu berücksichtigen, dass die Resultate sich oft eher über mehrere vermutete Zwischenschritte durch das Training ändern sollten, und somit insgesamt einen längeren Zeithorizont als den in diesem Abschnitt vorgestellten T1-T2 Vergleich. Ob sich andere Zusammenhänge und Einflussfaktoren zeigen, wird im Folgenden vorgestellt.

Für die *kognitive und emotionale Irritation* lassen sich neben den Mittelwerten auch Normwerte betrachten, die im Falle der kognitiven Irritation altersadjustiert sind (Mohr, Müller & Rigotti, 2005) und ebenfalls analysiert werden. Insgesamt liegen die Irritationswerte auf einem niedrigen, unauffälligen Niveau. Es findet sich lediglich ein einfacher Haupteffekt zwischen den Gruppen 1 und 2 in der kognitiven Irritation. Für die Kontrollvariablen zeigt sich zusätzlich ein Interaktionseffekt zwischen Geschlecht und Zeit auf die emotionale Irritation ( $F_{(1,159)}=3.70$ ,  $p=.056$ ). Dieser zeigt ein Sinken der emotionalen Irritation der weiblichen Probanden der Gruppen 1 und 2, wohingegen die Frauen der Gruppe 3 und die männlichen Probanden auf dem gleichen Niveau bleiben. Dieser Effekt ist unabhängig von Training.

Tabelle 30: Haupteffekte der Evaluationsebene „Resultate“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Gruppe 3 (c)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
			M	SE	M	SE	M	SE	F	p	F	p	F	p
Längerfristige Beanspruchungsfolgen – Irritation														
Kognitive Irritation	164	T1	2.96	0.15	2.61	0.14	2.83	0.25	0.01	.908	1.87	.158	0.01	.989
		T2	2.71 <sup>(b)</sup>	0.14	2.34 <sup>(a)</sup>	0.13	2.58	0.23						
Emotionale Irritation	164	T1	2.66	0.12	2.51	0.12	2.52	0.20	1.99	.160	0.78	.462	0.31	.731
		T2	2.56	0.13	2.31	0.12	2.44	0.21						
Irritation gesamt	164	T1	2.77	0.12	2.55	0.11	2.63	0.19	0.86	.354	1.36	.258	0.20	.823
		T2	2.62	0.12	2.32	0.12	2.49	0.20						
Längerfristige Beanspruchungsfolgen – körperliche Beschwerden														
GBB Erschöpfung	159	T1	5.35	0.48	4.73	0.45	4.13	0.78	0.09	.760	0.79	.454	0.37	.691
		T2	5.45	0.56	4.70	0.52	4.71	0.90						
GBB Magenbeschwerden	162	T1	2.61	0.43	2.80	0.40	2.16	0.68	1.62	.205	0.18	.835	0.92	.402
		T2	2.79	0.46	2.45	0.43	2.31	0.73						
GBB Gliederschmerzen	163	T1	6.14	0.53	5.24	0.49	6.31	0.84	1.47	.227	1.36	.260	0.35	.706
		T2	5.81	0.58	4.76	0.54	6.36	0.92						
GBB Herzbeschwerden	161	T1	2.10	0.40	1.66	0.39	1.80	0.65	0.09	.760	0.86	.424	0.93	.396
		T2	2.22	0.44	1.47	0.43	0.97	0.71						
GBB Beschwerdedruck	153	T1	15.87	1.61	14.53	1.54	14.50	2.56	0.72	.396	0.38	.686	0.26	.770
		T2	16.12	1.83	13.62	1.74	14.45	2.91						
Längerfristige Beanspruchungsfolgen – Wohlbefinden und Burnout														
Wohlbefinden (WHO-5)	163	T1	3.15	0.11	3.30	0.10	3.21	0.18	0.03	.862	1.65	.195	1.37	.256
		T2	3.09 <sup>(b)</sup>	0.11	3.45 <sup>(a)</sup>	0.11	3.20	0.18						
Burnout (emotionale Erschöpfung)	163	T1	2.77	0.15	2.63	0.14	2.81	0.23	0.22	.638	0.61	.545	0.67	.513
		T2	2.70	0.15	2.65	0.14	3.01	0.24						
Burnout (Zynismus)	163	T1	2.67	0.14	2.46	0.13	2.77	0.22	0.15	.704	0.98	.376	0.03	.970
		T2	2.75	0.15	2.54	0.14	2.81	0.24						

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Gruppe 3 (c)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
			M	SE	M	SE	M	SE	F	p	F	p	F	p
Schichtbezogene Beanspruchung – Veränderung der Befindlichkeit über die Schichtmessungen														
Vitalität (BBS)	145	T1	-0.89	0.10	-0.89	0.10	/	/	2.21	.139	0.22	.640	0.47	.496
		T2	-0.44	0.06	-0.35	0.06	/	/						
Gleichgewicht (BBS)	145	T1	-0.38	0.07	-0.30	0.06	/	/	0.15	.698	0.08	.773	1.29	.259
		T2	-0.39	0.08	-0.42	0.07	/	/						
Extraversion (BBS)	145	T1	-0.34	0.09	-0.33	0.08	/	/	0.64	.425	0.19	.663	0.25	.619
		T2	-0.47	0.11	-0.38	0.11	/	/						
Vigilanz (BBS)	145	T1	-0.80	0.09	-0.63	0.09	/	/	0.48	.491	0.31	.578	2.14	.146
		T2	-0.72	0.11	-0.76	0.11	/	/						
Schichtbezogene Beanspruchung – Workload zum Ende der Schichtmessungen														
NASA geistige Anforderungen	140	T1	40.45	2.30	36.45	2.20	/	/	0.46	.497	3.62	.059 <sup>+</sup>	1.08	.300
		T2	38.04 <sup>(b)</sup>	2.39	30.95 <sup>(a)</sup>	2.29	/	/						
NASA körperliche Anforderungen	140	T1	37.76	2.14	33.89	2.05	/	/	0.01	.941	4.46	.037 <sup>*</sup>	1.68	.197
		T2	39.33 <sup>b</sup>	2.43	31.41 <sup>a</sup>	2.33	/	/						
NASA zeitliche Anforderungen	140	T1	45.22	2.15	41.85	2.06	/	/	2.19	.141	1.90	.170	0.13	.721
		T2	44.45	2.50	39.95	2.39	/	/						
NASA Leistung	140	T1	81.40	1.43	82.93	1.37	/	/	0.99	.323	0.88	.350	0.11	.746
		T2	79.39	1.71	81.55	1.63	/	/						
NASA Anstrengung	140	T1	34.93	2.18	33.40	2.08	/	/	1.28	.260	1.69	.195	1.86	.175
		T2	34.79	2.28	29.22	2.18	/	/						
NASA Frustration	140	T1	22.24	2.33	21.25	2.23	/	/	0.28	.596	0.06	.808	0.06	.799
		T2	22.26	2.09	21.86	2.01	/	/						
Schichtbezogene Beanspruchung – Müdigkeit über die Schichtmessungen														
Mittlere Müdigkeit (KSS)	144	T1	3.60	0.14	3.43	0.14	/	/	0.01	.944	0.58	.448	0.12	.725
		T2	3.43	0.15	3.32	0.15	/	/						
Mittlere Ermüdung (KSS)	136	T1	1.09	0.23	0.80	0.22	/	/	0.23	.629	0.37	.546	0.46	.501
		T2	1.25	0.23	1.20	0.22	/	/						

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Zeitpunkt	Gruppe 1 (a)		Gruppe 2 (b)		Gruppe 3 (c)		Zeiteffekt		Gruppen-effekt		Interaktions-effekt	
			M	SE	M	SE	M	SE	F	p	F	p	F	p
Schichtbezogene Beanspruchung – kardiologische Werte während der Schichtmessungen														
Mittlere Herzfrequenz (Schläge pro Minute)	108	T1	91.8	1.4	90.3	1.3	/	/	0.54	.465	1.39	.241	0.52	.473
		T2	91.5	1.3	88.9	1.3	/	/						
Maximale Herzfrequenz (Schläge pro Minute)	108	T1	140.9	2.0	139.6	2.0	/	/	0.21	.646	0.09	.766	0.12	.725
		T2	136.6	1.9	136.6	1.9	/	/						
Minimale Herzfrequenz (Schläge pro Minute)	107	T1	57.3	1.2	56.1	1.2	/	/	0.32	.576	0.85	.358	0.01	.919
		T2	57.6	1.2	56.2	1.2	/	/						
Typische Kennzahlen – Dauer der Einsatzfahrten														
„Hilfsfrist“ (Dauer der Einsatzfahrten)	134	T1	7.03	0.29	6.39	0.28	/	/	4.11	.045*	2.27	.135	0.17	.681
		T2	7.11	0.34	6.69	0.33	/	/						

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: N=Anzahl der Probanden, M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Gruppen

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Zur weiteren Illustrierung jenseits des Evaluationsdesigns werden die Normwerte der Skalen herangezogen und mit den entsprechenden Prozenträngen der Studienteilnehmer verglichen (vgl. Abbildung 30). Eine Person mit einem Prozentrang von 66 bedeutet, dass 65% der Normierungsstichprobe bessere Werte haben und 34% schlechtere. Eine „normale“ Stichprobe sollte demzufolge in den vier angegebenen Prozentrangbereichen jeweils 25% der Personen haben. Als kritisch werden Werte jenseits des 75. Prozentrangs angesehen. Die vorhandene Stichprobe zeigt ein insgesamt sehr positives Bild, denn besonders die Prozentränge bis 50% kommen deutlich häufiger als zu erwarten vor und die Prozentränge über 75% nur sehr selten. Dieses Ergebnis ist für die kognitive Irritation deutlich höher ausgeprägt als für die emotionale, in der die Verteilung schon eher dem Durchschnitt entspricht. Insgesamt ist die Irritation also sehr positiv ausgeprägt und nur wenige Probanden weisen kritische Werte auf. Trainingseffekte lassen sich jedoch keine finden.

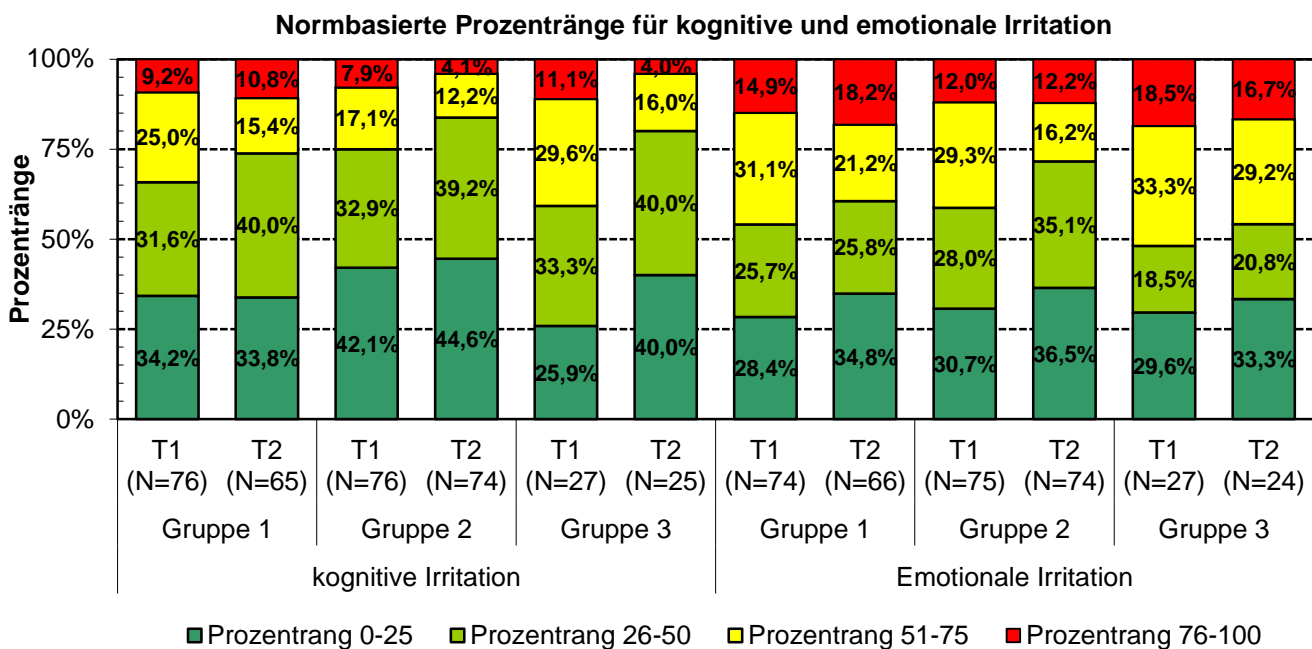


Abbildung 30: Prozentränge für kognitive und emotionale Irritation

Als längerfristige Beanspruchungsfolgen wurden auch *körperliche Beschwerden* in den Bereichen Erschöpfungsneigung, Magenbeschwerden, Gliederschmerzen und Herzbeschwerden mittels GBB (Brähler et al., 2008; vgl. Abschnitt 3.2.1.4) erfasst. Auch für diesen Bereich finden sich keinerlei kurzfristigen Effekte des Trainings, lediglich für die Magenbeschwerden gibt es einen signifikanten Einfluss des Alters dahingehen, dass die älteren Probanden deutlich höhere Werte aufweisen als die jüngeren ( $F_{(1,157)}=5.17, p=.024$ ). Dieser Effekt zeigt sich auch für den Gesamtwert (Beschwerdedruck) noch tendenziell ( $F_{(1,148)}=2.76, p=.099$ ), was allerdings aufgrund der gesundheitlichen Entwicklungen mit dem Alter und in einer erheblichen belastenden Tätigkeit wie dem Rettungsdienst nicht weiter verwundert. Ähnlich wie bei der Irritation können die Werte der Teilnehmer beim GBB mit geschlechts- und altersadjustierten bevölkerungsrepräsentativen Prozenträngen verglichen werden und werden in den folgenden Abbildung 31 und Abbildung 32 genutzt, um unabhängig vom Evaluationsdesign die Beschwerden der Studienteilnehmer zu illustrieren.



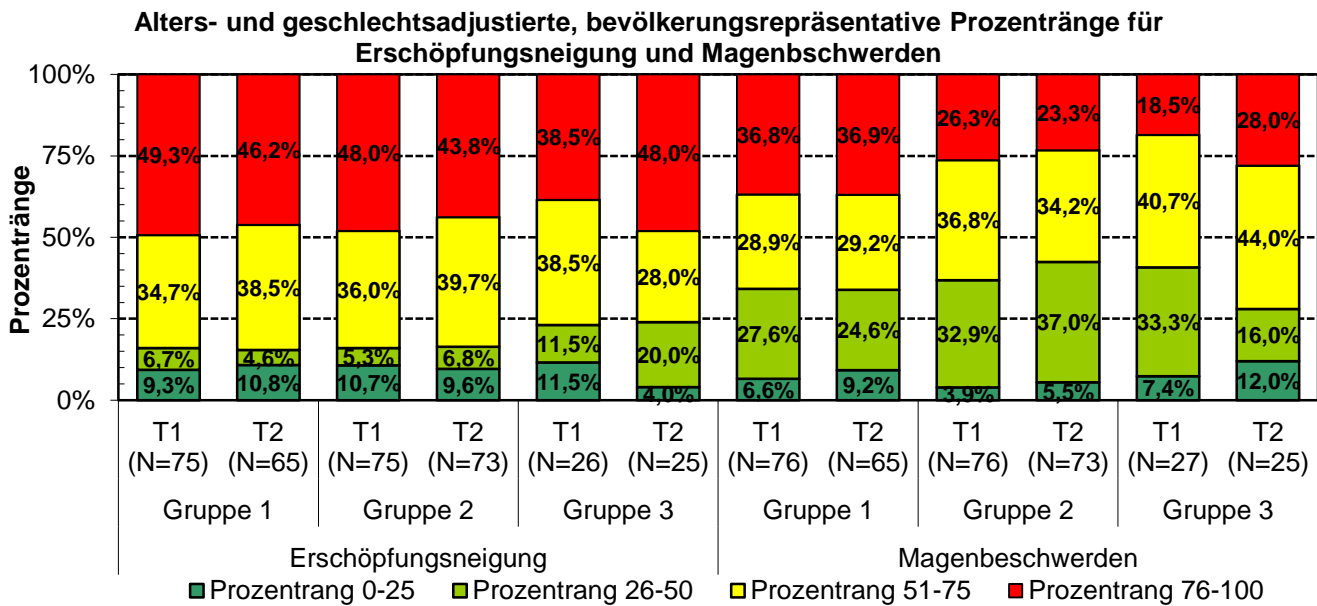


Abbildung 31: Erschöpfungsneigung und Magenbeschwerden der Studiengruppen im Vergleich zur Allgemeinbevölkerung

Auffällig ist im Gegensatz zur Irritation, dass bei den körperlichen Beschwerden die kritischen Werte dominieren und die Studienteilnehmer insgesamt erheblich mehr körperliche Beschwerden angeben als die Allgemeinbevölkerung. Besonders ausgeprägt ist dies bei der Erschöpfungsneigung und den Gliederschmerzen. Relativ deutlich illustrieren die Abbildungen auch, dass es kaum oder keine Veränderungen im Zeitverlauf gibt.

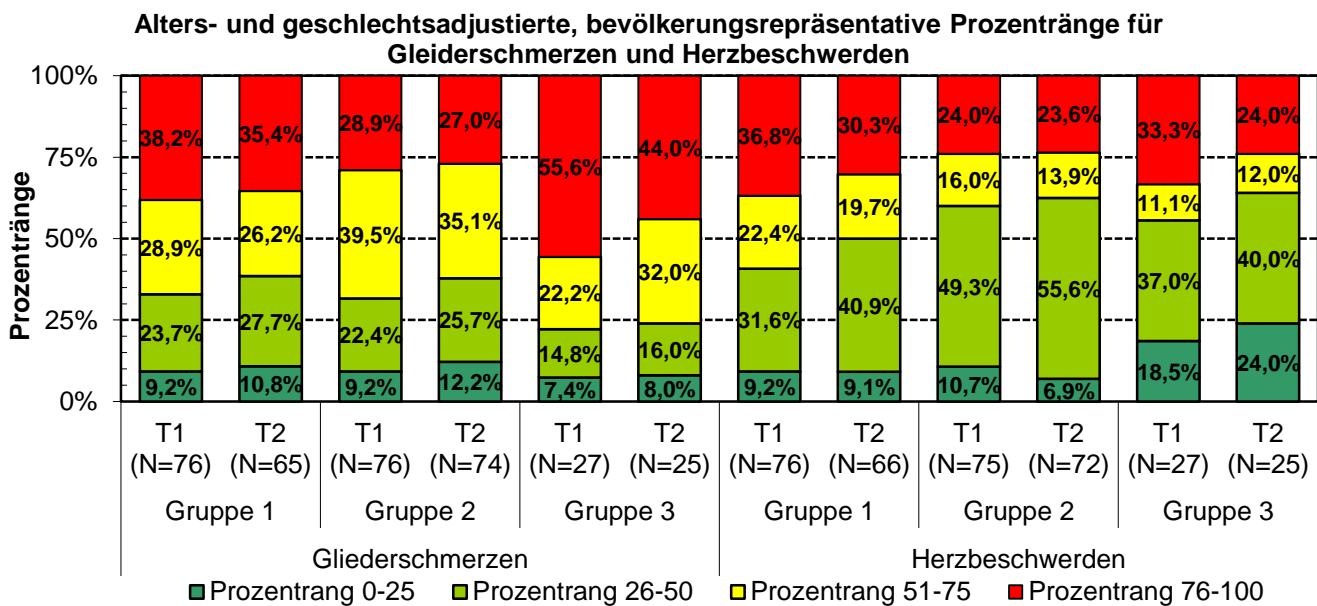


Abbildung 32: Gliederschmerzen und Herzbeschwerden der Studiengruppen im Vergleich zur Allgemeinbevölkerung

Als weitere längerfristigen Beanspruchungsfolgen wurden *Burnout und Depression/Wohlbefinden* der Teilnehmer erfasst (vgl. Abschnitt 3.2.1.4). Für das Wohlbefinden und die beiden Burnoutdimensionen emotionale Erschöpfung und Zynismus können erneut keinerlei Effekte des Trainings festgestellt werden. Ein

minimaler Einzeltrend weist darauf hin, dass Gruppe 2 ein höheres Wohlbefinden aufweist als Gruppe 1, und es zeigen sich – unabhängig von den Studiengruppen und der Zeit – Alters- und Geschlechtseffekte auf die Burnoutkomponente Zynismus. Abbildung 33 illustriert die beiden Effekte. Dort ist zu erkennen, dass Männer und jüngere Probanden insgesamt höher Werte im Zynismus aufweisen als Frauen und ältere Probanden. Das Geschlecht ( $F_{(1,158)}=8.27, p=.005$ ) spielt dabei eine größere Rolle, als das Alter ( $F_{(1,158)}=4.85, p=.029$ ).

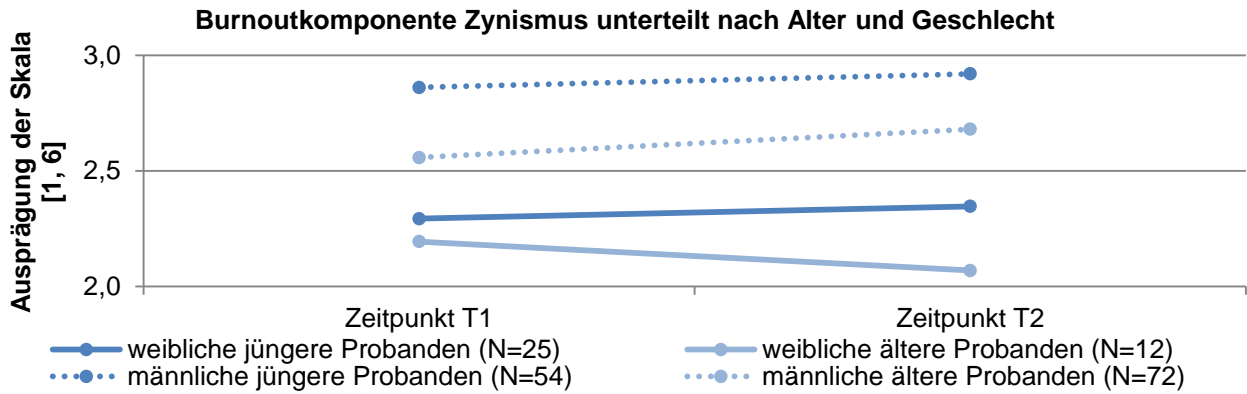


Abbildung 33: Zynismus (Burnout) in Abhängigkeit von Alter und Geschlecht

Wie bei den anderen längerfristigen Beanspruchungsfolgen, wird im Folgenden kurz deskriptiv die Häufigkeit von Depressionen in der Stichprobe mittels des Screeningwerts des WHO-5 (vgl. Abschnitt 3.2.1.4) dargestellt. Das Screening liefert eine Verdachtsdiagnose, von der 34% (positiver Prädiktionswert) dann tatsächlich an einer Depression leiden. Abbildung 34 zeigt die Verteilung – der Anteil von Probanden, die zum Messzeitpunkt vermutlich tatsächlich an einer Depression leiden, liegt im Mittel von 7.0% (zwischen 4.1 und 8.5%) und ist somit vergleichbar zur Gesamtbevölkerung, in der die Punktprävalenz für Depression bei ca. 7% (2-3% der Männer und 5-9% der Frauen) liegt (Wittchen et al., 2011). Da die Stichprobe jedoch zum großen Teil aus männlichen Probanden besteht, sind diese Wert eher kritisch zu sehen.

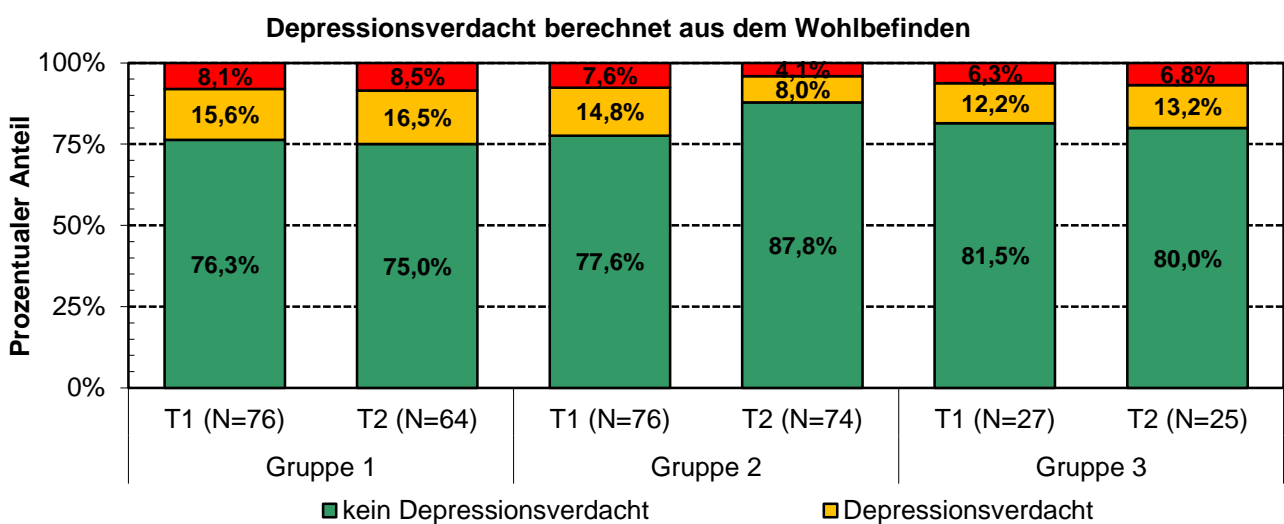


Abbildung 34: Depressionsscreening, ca. 34% der Personen mit Depressionsverdacht weisen tatsächlich eine Depression auf

Neben den längerfristigen Beanspruchungsfolgen wurden während der erhobenen Schichten auch unmittelbar *schichtbezogene Beanspruchungen* erfasst (vgl. Abschnitt 3.2.1.4 zur Vorstellung der Verfahren). Im Bereich der Befindlichkeit wurden Vitalität, Gleichgewicht, Extraversion und Vigilanz mehrfach erfasst. Da jeweils unterschiedlich viele Messungen vorliegen (vgl. Abschnitt 3.4.3), wurde der Mittelwert der Differenzen zwischen Schichtbeginn und -ende der einzelnen erfassten Schichten genutzt. Die Daten zeigen jeweils, dass alle vier Befindlichkeitsskalen über die Schichten hinweg abnehmen. Es gibt jedoch keinerlei Effekte zwischen den Zeitpunkten oder Gruppen und auch keine Interaktionseffekte zwischen den Gruppen und Zeitpunkten. Das Training hat demzufolge keinerlei Einfluss auf die Befindlichkeitsänderung über die Schichten, allerdings tritt eine Effekt des Geschlechts auf die Entwicklung des intrapsychischen Gleichgewicht über die Schicht auf ( $F_{(1,141)}=5.26, p=.023$ ), der auch für die anderen Befindlichkeitsskalen ähnlich gelagert ist (vgl. Abbildung 35): Männer empfinden eine höhere Verringerung des Gleichgewichtszustandes als Frauen über die Schichten.

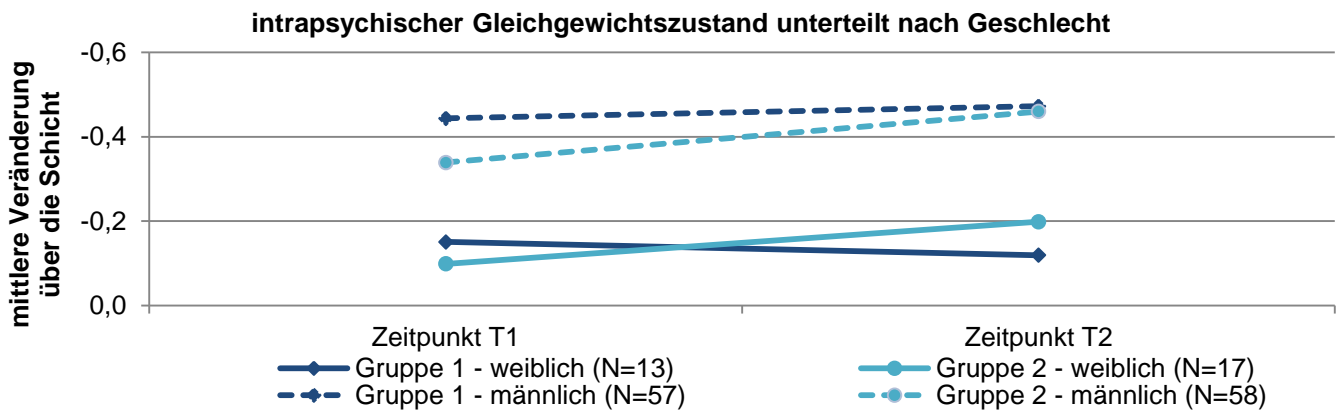


Abbildung 35: Mittlere Veränderung des intrapsychischen Gleichgewichtszustands über die Rettungsdienstschichten nach Geschlecht

Als weitere schichtbezogene Beanspruchung wurde der Workload mit dem NASA-TLX erfasst (vgl. Abschnitt 3.2.1.4). Effekte, die auf die Wirksamkeit des Trainings hinweisen, finden sich nicht, allerdings gibt es für die Bereiche geistige und körperliche Anforderungen signifikante Gruppenunterschiede, die im Einzelnen zwischen der Gruppen 1 und 2 (vgl. Abbildung 36) liegen.

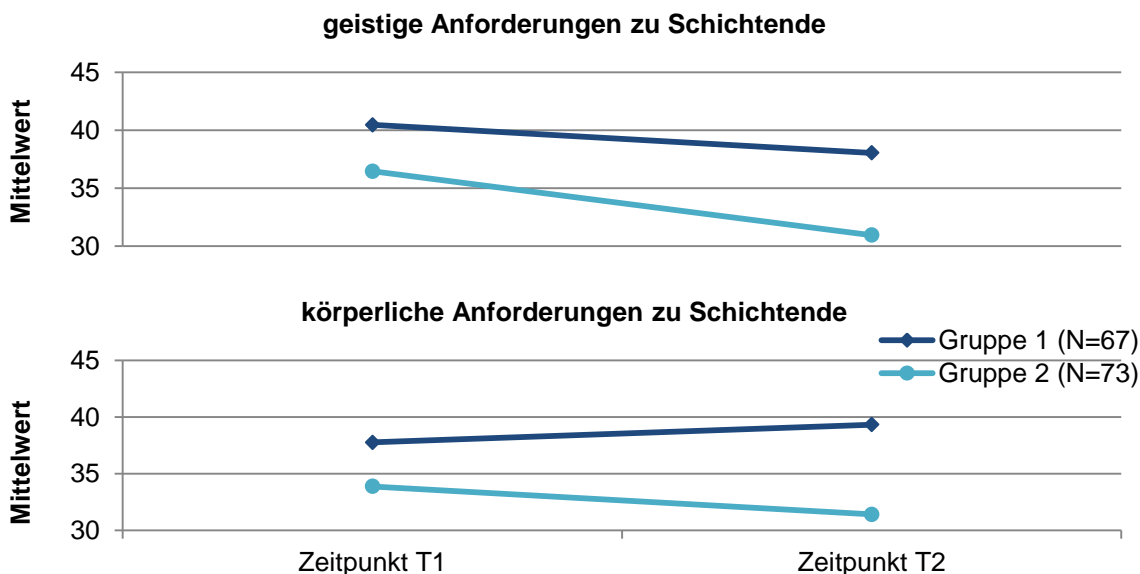


Abbildung 36: Geistige und körperliche Anforderungen zu Schichtende

In beiden Fällen bewerten die Probanden der Interventionsgruppe bereits zu T1 die Schichten mit höheren Anforderungen als die Wartegruppe. Die zeitlichen Anforderungen weisen sowohl einen signifikanten Alters- ( $F_{(1,136)}=6.31$ ,  $p=.013$ ) als auch Geschlechtseffekt ( $F_{(1,136)}=4.26$ ,  $p=.041$ ) auf. Er zeigt sich darin, dass die älteren Probanden und die Frauen höhere zeitliche Anforderungen wahrnehmen als die jüngeren bzw. die Männer. Tendenziell findet sich auch ein Interaktionseffekt zwischen den Messzeitpunkten und dem Geschlecht auf die Bewertung der Frustration zum Schichtende ( $F_{(1,136)}=3.75$ ,  $p=.055$ ). Dieser zeigt eine sinkende Bewertung der Frustration bei den weiblichen Probanden der Interventionsgruppe 1. Alle anderen Teilgruppen bleiben bei ihren Bewertungen gleich. Inwieweit dieser Befund auf differenzielle Wirksamkeit der Trainings für Männer und Frauen hinweist, ist unklar, da gerade auf der vierten Ebene viele weitere Einflussfaktoren eine Auswirkung haben können.

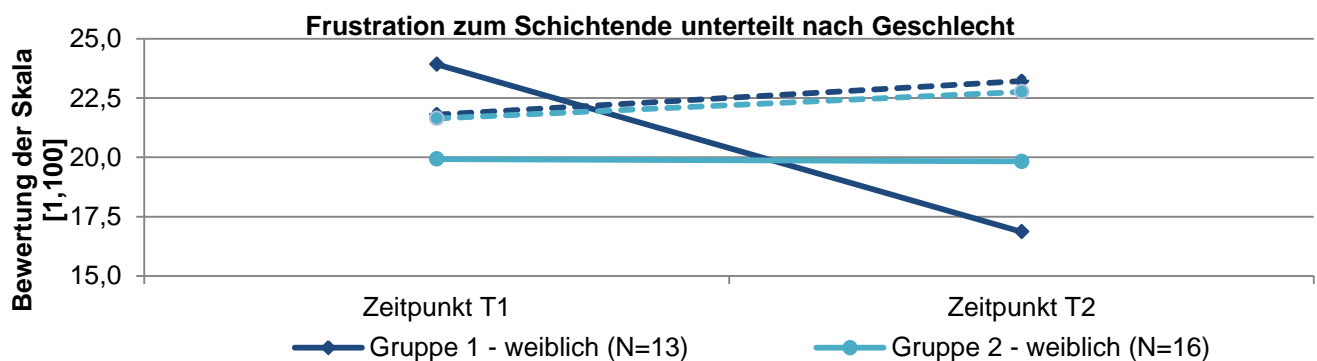


Abbildung 37: Frustration zu Schichtende in Abhängigkeit vom Geschlecht

Müdigkeit und Ermüdung über die Schicht wurde als weiterer Parameter schichtbezogener Beanspruchung mehrfach während der Schichten mittels KSS (Åkerstedt & Gillberg, 1990) erhoben. Auch hier fanden sich viele fehlende Werte (vgl. Abschnitt 3.2.4), so dass in die Auswertungen zum einen die Müdigkeit am Schichtende und zum anderen die mittlere Ermüdung über die Schichten, also die Differenz zwischen Schichtende und Schichtbeginn, aufgenommen wurden. Insgesamt kann für Müdigkeit oder Ermüdung kein Einfluss des Trainings gefunden werden. Es findet sich allerdings ein signifikanter Einfluss des Geschlechts auf die Müdigkeit. Frauen geben insgesamt höhere Müdigkeitswerte an als Männer ( $F_{(1,136)}=5.14$ ,  $p=.025$ ).

Als letzter Bereich der schichtbezogenen Beanspruchung wurden während einzelner Schichten EKGs durchgeführt. Für diese kardiologischen Parameter konnten aufgrund einer Reihe von Messproblemen (vgl. Abschnitt 3.2.4) nur für 108 Probanden Auswertungen vorgenommen werden. Weder für die mittlere, maximale oder minimale Herzfrequenz zeigen sich Effekte des Trainings, allerdings ist auffällig, dass bei allen Probanden die Herzfrequenzwerte relativ hoch sind. Abbildung 38 verdeutlicht die Werte: Deutlich zeigen sich die fehlenden Gruppen- und Zeiteffekte, klar wird aber auch, dass die minimalen Werte in einem gesunden Bereich liegen, während die mittlere Herzfrequenz insgesamt (zu) hoch ist. Die maximalen Werte liegen bei ca. 140 Schlägen pro Minute, gehen jedoch bis zu annähernd 200 Schlägen hoch. Diese Werte sind eher bedenklich. Von den 135 EKGs, die insgesamt zum Zeitpunkt T1 ausgewertet werden konnten, wiesen 47 eine erhöhte Grundrate auf und weitere 17 Probanden erhielten die ärztliche Empfehlung, ihre Werte bei einem niedergelassenen Arzt oder Kardiologen genauer abklären zu lassen.

Die minimale Herzfrequenz ist die einzige Größe, die einen Effekt mit dem Alter aufweist und zwar sowohl einen signifikanten Haupt- ( $F_{(1,103)}=4.87$ ,  $p=.030$ ) als auch tendenziellen Interaktionseffekt ( $F_{(1,103)}=3.47$ ,  $p=.066$ ). Ältere Probanden weisen eine insgesamt höhere minimale Herzfrequenz auf, die über die Zeit relativ stabil bleibt. Die minimale Herzfrequenz der jüngeren Probanden steigt zum Zeitpunkt T2 etwas an.

Es gibt jedoch auch bei dieser Auswertung keinerlei Trainingseffekt, beide Gruppen entwickeln sich gleichermaßen.

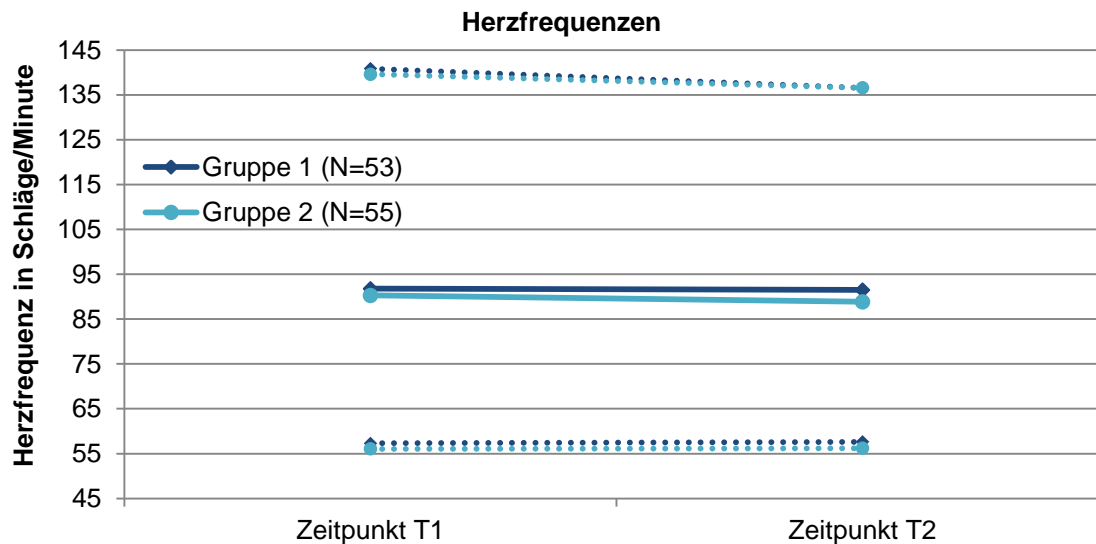


Abbildung 38: mittlere (durchgezogene Linie), maximale und minimale (gepunktete Linie) Herzfrequenz in den Versuchsgruppen in den Schichtmessungen

Als letzter Parameter der Resultateebene wurde als *typische Kennzahl* im Rettungsdienst die „Hilfsfrist“ betrachtet. Wie bereits in Abschnitt 2 dargestellt, ist die sogenannte Hilfsfrist ein wichtiger planerischer Kennwert im Rettungsdienst und die hier verwendete „Dauer der Einsatzfahrten“ entspricht als Proxy nur einem Teil der Hilfsfrist. Ziel der Trainings ist, dass die Einsatzfahrer sicherer am Einsatzort ankommen, dennoch wäre es wirtschaftlich, wenn sich die Dauer der Einsatzfahrten dadurch nicht unnötig erhöht. Daher wurde mit der Dauer der zusammengefassten RTW- und NEF-Einsatzfahrten zum Einsatzort die tatsächliche Länge der Einsatzfahrten erfasst.

Für diese „Hilfsfrist“ gibt einen signifikanten Haupteffekt der Zeit, dahingehend, dass die Einsatzfahrten zum Zeitpunkt T2 etwas länger werden im Vergleich zu T1. Die Veränderung ist eher gering: Die Interventionsgruppe 1 hat zu T2 5 Sekunden längere Fahrten, die Wartegruppe 2 fährt 19 Sekunden länger als zu T1. Die Einsatzfahrten verändern sich in den Gruppen also gleichläufig und eher entgegen den Erwartungen.

Es gibt für die Hilfsfristen neben dem Zeiteffekt auch erneut Effekte der Kontrollvariablen - einen tendenziellen Haupteffekt des Geschlechts ( $F_{(1,136)}=3.18, p=.077$ ) und einen tendenziellen Interaktionseffekt mit dem Alter der Einsatzfahrer ( $F_{(1,136)}=3.14, p=.079$ ). Zum Zeitpunkt T1 sind alle Gruppen auf dem gleichen Niveau, die Hilfsfrist der weiblichen älteren Probanden steigt zu T2 an, die der männlichen älteren Probanden nimmt eher ab. Diese Effekte treten unabhängig vom Training.

#### 4.2.3 Prüfung der Effekte der Messungen

Die im Studiendesign eingeführte Gruppe 3 diente zur Kontrolle, ob anstatt bzw. zusätzlich zum Training Effekte auftreten, die ausschließlich darauf zurückzuführen sind, dass mit den Teilnehmern bereits Messungen durchgeführt wurden. Gruppe 3 wurde vor dem Training nur mit Fragebögen befragt, nicht jedoch bei ihren Einsatzfahrten und in den Schichten gemessen. Vor diesem Hintergrund wird im Folgenden untersucht, ob es Variablen gibt, in denen sich Gruppe 1 von Gruppe 3 zum Zeitpunkt T2 unterscheidet. Tabelle 31 enthält die Ergebnisse einfaktorierlicher Varianzanalysen für alle getesteten Variablen zum Zeitpunkt T2 zwischen den beiden Gruppen.

Tabelle 31: Kontrolle der Messeffekte: Unterschiede zwischen Gruppe 1 und 3 zum Zeitpunkt T2

Abhängige Variable zum Zeitpunkt T2	N <sub>Gruppe1</sub> / N <sub>Gruppe3</sub>	Gruppe 1		Gruppe 3		Gruppen- effekt	
		M	SD	M	SD	F	p
Wissen	71/25	9.94	2.65	9.55	3.36	0.40	.528
Extraversion	66/25	1.58	0.44	1.63	0.46	0.45	.504
Sensation Seeking	66/25	2.31	0.44	2.30	0.49	0.00	.992
Gewissenhaftigkeit	66/25	2.67	0.52	2.59	0.63	0.73	.396
Unverträglichkeit	66/25	2.19	0.42	2.24	0.51	0.58	.448
Reaktanz	66/25	1.46	0.44	1.53	0.43	0.90	.346
Risikowahrnehmung bei schweren Regelverstößen	66/25	4.51	0.36	4.53	0.39	0.09	.765
Risikowahrnehmung bei mittelschweren Regelverstößen	66/25	3.96	0.48	3.99	0.64	0.00	.992
Risikowahrnehmung bei leichten Regelverstößen	66/25	3.56	0.50	3.62	0.54	0.15	.700
Risikowahrnehmung bei normalem Fahren	66/25	2.98	0.51	3.09	0.58	0.57	.453
Risk Seeking	66/25	1.76	0.72	2.06	0.97	3.74	.056 <sup>+</sup>
Fahrtkompetenz allgemein	66/25	7.10	0.99	7.40	1.02	1.83	.179
Fahrtkompetenz Rettungsdienst	66/25	6.74	1.49	7.20	1.34	2.02	.159
Unfallwahrscheinlichkeit allgemein	66/25	4.68	1.85	4.56	1.66	0.23	.637
Unfallwahrscheinlichkeit Rettungsdienst	66/25	4.33	1.76	5.40	2.18	5.33	.023 <sup>*</sup>
Durchschnittliche Geschwindigkeit (in km/h)	60/18	53.34	11.50	48.11	10.28	2.50	.118
Maximale Geschwindigkeit (in km/h)	60/18	112.06	21.35	106.97	21.23	1.01	.318
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	60/18	2.58	1.04	3.39	1.05	8.16	.006 <sup>**</sup>
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	60/18	0.27	0.05	0.28	0.04	0.31	.580
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	60/18	0.00	0.00	0.00	0.00	0.19	.661
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	60/18	0.00	0.00	0.00	0.00	0.19	.661
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	60/18	-3.03	1.18	-3.25	0.71	0.27	.606
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	60/18	-0.29	0.06	-0.30	0.05	0.66	.420
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	60/18	0.14	0.45	0.14	0.22	0.01	.937
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	60/18	0.01	0.03	0.02	0.03	1.18	.281
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	60/18	0.00	0.02	0.00	0.00	1.38	.244
Anzahl Trigger pro Minute negative Längsbeschleunigung (5,9 m/s <sup>2</sup> )	60/18	0.00	0.01	0.00	0.00	0.76	.387
Maximale positive Querbeschleunigung (in m/s <sup>2</sup> )	60/18	3.97	1.07	4.29	0.70	0.82	.369
Mittlere positive Querbeschleunigung (in m/s <sup>2</sup> )	60/18	0.41	0.13	0.41	0.14	0.01	.945
Maximale negative Querbeschleunigung (in m/s <sup>2</sup> )	60/18	-3.92	1.34	-3.85	1.24	0.02	.889
Mittlere negative Querbeschleunigung (in m/s <sup>2</sup> )	60/18	-0.39	0.12	-0.37	0.12	0.10	.757

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable zum Zeitpunkt T2	N <sub>Gruppe1</sub> / N <sub>Gruppe3</sub>	Gruppe 1		Gruppe 3		Gruppen- effekt	
		M	SD	M	SD	F	p
Dauer Trigger Querschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	60/18	0.00	0.01	0.01	0.05	1.54	.219
Anzahl Trigger pro Minute Querschleunigung (6,9 m/s <sup>2</sup> )	60/18	0.00	0.00	0.00	0.00	0.47	.494
Fahrerbeurteilung sichere Fahrweise	70/25	79.39	14.32	78.02	15.10	0.07	.791
Fahrerbeurteilung vorausschauende Fahrweise	70/25	81.04	12.35	82.67	10.11	0.36	.551
Fahrerbeurteilung rasante Fahrweise	70/25	41.57	22.08	43.19	22.32	0.13	.721
Fahrerbeurteilung hektische Fahrweise	70/25	20.38	16.16	18.81	13.50	0.11	.744
Beifahrerbeurteilung sichere Fahrweise	63/23	88.68	13.43	85.47	11.96	1.12	.293
Beifahrerbeurteilung vorausschauende Fahrweise	63/23	87.80	13.37	85.22	11.35	0.76	.387
Beifahrerbeurteilung rasante Fahrweise	63/23	36.59	23.43	32.57	24.75	0.37	.547
Beifahrerbeurteilung hektische Fahrweise	63/23	17.26	20.55	13.08	12.68	0.78	.381
Kognitive Irritation	66/25	2.69	1.26	2.60	1.11	0.10	.751
Emotionale Irritation	66/25	2.55	1.11	2.44	1.04	0.13	.719
Irritation gesamt	66/25	2.60	1.09	2.50	0.96	0.14	.711
GBB Erschöpfung	65/25	5.34	4.86	4.80	4.14	0.32	.576
GBB Magenbeschwerden	65/25	2.85	3.67	2.44	3.25	0.30	.586
GBB Gliederschmerzen	65/25	5.85	5.23	6.52	4.18	0.23	.633
GBB Herzbeschwerden	66/25	2.23	4.14	1.08	1.89	2.03	.158
GBB Beschwerdedruck	63/25	16.03	15.56	14.84	11.48	0.19	.668
Wohlbefinden (WHO-5)	65/25	3.10	1.00	3.19	0.97	0.15	.700
Burnout (emotionale Erschöpfung)	65/25	2.70	1.29	3.03	1.12	1.17	.283
Burnout (Zynismus)	65/25	2.71	1.23	2.74	1.31	0.11	.736
Vitalität (BBS)	70/25	-0.45	0.50	-0.44	0.44	0.00	.995
Gleichgewicht (BBS)	70/25	-0.41	0.71	-0.42	0.60	0.05	.818
Extraversion (BBS)	70/25	-0.48	0.94	-0.36	1.03	0.19	.662
Vigilanz (BBS)	70/25	-0.73	0.96	-0.90	0.77	0.94	.335
NASA geistige Anforderungen	68/24	38.23	20.43	38.61	21.20	0.00	.948
NASA körperliche Anforderungen	68/24	39.09	19.82	39.90	19.19	0.00	.953
NASA zeitliche Anforderungen	68/24	44.86	20.90	46.57	19.26	0.07	.791
NASA Leistung	68/24	79.65	13.49	82.97	10.42	1.00	.320
NASA Anstrengung	68/24	34.67	19.34	35.30	15.68	0.01	.927
NASA Frustration	68/24	22.15	16.63	22.41	19.75	0.09	.766
Mittlere Müdigkeit (KSS)	71/25	3.39	1.23	3.36	1.08	0.08	.774
Mittlere Ermüdung (KSS)	68/23	1.31	2.10	1.23	2.10	0.02	.895
Mittlere Herzfrequenz (Schläge pro Minute)	54/22	91.04	10.22	84.86	10.99	5.64	.020*
Maximale Herzfrequenz (Schläge pro Minute)	54/22	136.41	14.12	135.68	8.44	0.01	.945
Minimale Herzfrequenz (Schläge pro Minute)	54/22	57.50	9.21	54.14	9.10	2.40	.126
„Hilfsfrist“ (Dauer der Einsatzfahrten, in mm:ss)	65/22	7.05	3.05	5.98	1.56	2.68	.105

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht); Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

In fast allen gemessenen Variablen gibt es keine Unterschiede zwischen Gruppe 1 und 3 zum Zeitpunkt T2. Lediglich vier Variablen weisen signifikante Unterschiede auf. Risk Seeking ist in der Gruppe 3 etwas höher ausgeprägt. Ebenso schätzen die Probanden der Gruppe 3 ihre eigene Unfallwahrscheinlichkeit höher ein als die der Gruppe 1. Die mittlere Herzfrequenz ist bei der Gruppe 3 geringer als bei der Gruppe 1 und die maximale Längsbeschleunigung ist in der Gruppe 3 mit dem deutlichsten Effekt höher als in Gruppe 1. Obwohl mit großer Vorsicht zu interpretieren, da nur wenige Unterschiede bei einer Vielzahl von Tests auftreten, könnten kleinere Effekte des Messens vorliegen. So wäre denkbar, dass die niedrigere Maximalbeschleunigung in Gruppe 1 auf einen „disziplinierenden“ Effekt der Messung hinweist, während die höhere mittlere Herzfrequenz in dieser Gruppe aus einer gewissen „Störung“ durch die Messung resultieren könnte. Für risk seeking und Einschätzung der eigenen Unfallwahrscheinlichkeit finden sich auch in der Hauptauswertung (vgl. Abschnitt 4.2.2.1) Gruppenunterschiede, aber keine Effekte des Trainings; bei der Maximalbeschleunigung und der mittleren Herzfrequenz finden sich in den Hauptauswertungen (Vgl. Abschnitt 4.2.2.2 und 4.2.2.3) keinerlei Effekte der Gruppen oder des Trainings, so dass nicht von systematischen Verzerrungen durch Messeffekte auszugehen ist.

#### 4.2.4 Prüfung der Voraussetzung zur Zusammenführung der Messzeitpunkte

Aufgrund des komplexen Designs liegen nur für Gruppe 2 (die Wartegruppe) Daten zum dritten Messzeitpunkt vor. In diesem Abschnitt soll nun geprüft werden, ob die Messungen der Wartegruppe verschoben werden können, um im weiteren Verlauf zu besserer Verständlichkeit und zur Erhöhung der statistischen Power alle 3 Gruppen jeweils mit einer Messung vor und einer nach dem Training vergleichen zu können. Damit könnten Einflussfaktoren auf den Trainingserfolg über die gesamte Stichprobe ermittelt werden, da alle Teilnehmer am Training teilgenommen haben. Um eine Verschiebung durchführen zu können, muss mindestens eine der folgenden Voraussetzungen erfüllt sein (vgl. Abbildung 39 zur Veranschaulichung):

1. Ausprägung der Gruppe 2 zum Messzeitpunkt T2 unterscheidet sich nicht von der Ausprägung der Gruppe 1 zum Messzeitpunkt 1
2. Ausprägung in der Gruppe 2 verändert sich nicht im Verlauf zwischen den Messzeitpunkten T1 und T2

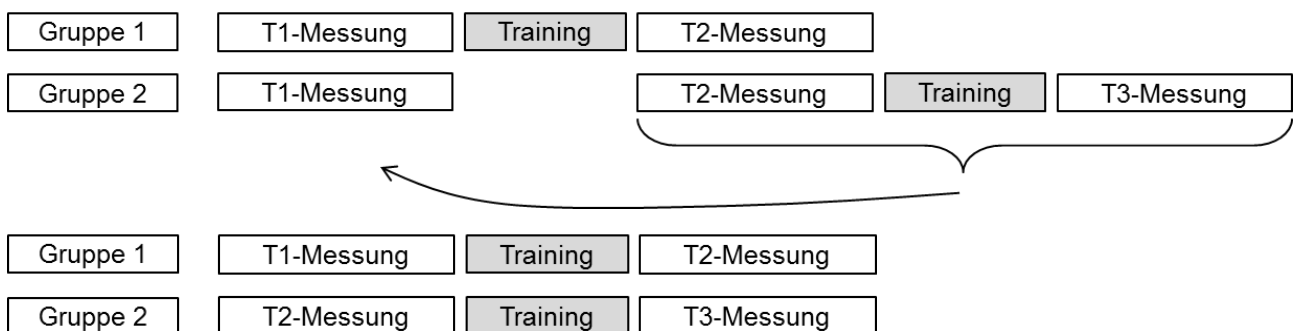


Abbildung 39: Veranschaulichung der Zusammenführung der Messzeitpunkte

Tabelle 32 zeigt für alle abhängigen Variablen, ob jeweils die erste bzw. zweite Bedingung zur Verschiebung der Daten erfüllt ist. Dabei werden bereits Trends, also Unterschiede mit einer 10% Irrtumswahrscheinlichkeit, als Voraussetzungsverletzung angesehen. Von den insgesamt 67 Variablen können die Voraussetzungen zur Verschiebung der Daten bei 16 Variablen überhaupt nicht erfüllt werden. Die meisten Variablen kommen dabei aus den Verhaltens- und Resultatevariablen. Bei 21 Variablen kann zumindest



eine der Voraussetzungen erfüllt werden. Beide Voraussetzungen werden bei 30 Variablen erfüllt. Dennoch weisen zu viele Variablen zu große Unterschiede zwischen den Messungen auf, so dass in den weiteren Auswertungen die Datenstruktur beibehalten wird und keine Verschiebung der Daten vorgenommen wird.

Tabelle 32: Bestimmung der Voraussetzungen zur Verschiebung der Messzeitpunkte der Gruppe 2

Abhängige Variable	Voraussetzung 1: Unterschied T2 Gruppe 2 zu T1 Gruppe 1			Voraussetzung 2: Veränderung über die Zeit T1/T2 in Gruppe 2		
	T-Wert	p-Wert	erfüllt	T-Wert	p-Wert	erfüllt
<b>Lernebene</b>						
Extraversion	0.438	.662	ja	0.573	.569	ja
Sensation Seeking	0.337	.737	ja	0.590	.557	ja
Gewissenhaftigkeit	1.056	.293	ja	1.186	.239	ja
(Un-)Verträglichkeit	0.036	.971	ja	1.396	.167	ja
Reaktanz	-0.168	.867	ja	0.075	.940	ja
Risikowahrnehmung bei schweren Regelverstößen	-0.135	.893	ja	-0.321	.749	ja
Risikowahrnehmung bei mittelschweren Regelverstößen	1.456	.148	ja	-1.565	.122	ja
Risikowahrnehmung bei leichten Regelverstößen	1.680	.095	nein	-0.022	.983	ja
Risikowahrnehmung bei normalem Fahren	1.360	.176	ja	0.308	.759	ja
Risk Seeking	1.624	.106	ja	2.708	.008	nein
Fahrtkompetenz allgemein	-2.129	.035	nein	1.018	.312	ja
Fahrtkompetenz Rettungsdienst	-0.935	.351	ja	1.870	.065	nein
Unfallwahrscheinlichkeit allgemein	-2.449	.015	nein	-.571	.570	ja
Unfallwahrscheinlichkeit Rettungsdienst	-1.661	.099	nein	-1.059	.293	ja
<b>Verhaltensebene*</b>						
Durchschnittliche Geschwindigkeit (in km/h)	1.820	.071	nein	-1.108	.272	ja
Maximale Geschwindigkeit (in km/h)	0.732	.465	ja	-0.195	.846	ja
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	2.260	.026	nein	1.457	.150	ja
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	2.113	.037	nein	2.776	.007	nein
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	1.689	.096	nein	1.564	.123	ja
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	1.693	.095	nein	1.544	.128	ja
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	-0.723	.471	ja	-2.045	.045	nein
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	-1.631	.106	ja	-2.788	.007	nein
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	2.312	.024	nein	2.206	.031	nein
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	1.795	.077	nein	2.133	.037	nein
Dauer Trigger negative Längsbeschleunigung (in 0.05 sek/min) (5,9 m/s <sup>2</sup> )	1.768	.081	nein	1.620	.111	ja
Anzahl Trigger pro Minute negative Längsbeschleunigung (5,9 m/s <sup>2</sup> )	1.612	.112	ja	1.659	.102	ja

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	Voraussetzung 1: Unterschied T2 Gruppe 2 zu T1 Gruppe 1			Voraussetzung 2: Veränderung über die Zeit T1/T2 in Gruppe 2		
	T-Wert	p-Wert	erfüllt	T-Wert	p-Wert	erfüllt
Maximale positive Querbeschleunigung (in m/s <sup>2</sup> )	3.142	.002	nein	2.633	.011	nein
Mittlere positive Querbeschleunigung (in m/s <sup>2</sup> )	2.659	.009	nein	3.094	.003	nein
Maximale negative Querbeschleunigung (in m/s <sup>2</sup> )	-2.271	.026	nein	-2.267	.027	nein
Mittlere negative Querbeschleunigung (in m/s <sup>2</sup> )	-2.134	.036	nein	-2.338	.023	nein
Dauer Trigger Querbeschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	2.255	.027	nein	2.188	.033	nein
Anzahl Trigger pro Minute Querbeschleunigung (6,9 m/s <sup>2</sup> )	2.153	.035	nein	2.256	.028	nein
Fahrerbeurteilung sichere Fahrweise	0.937	.350	ja	2.492	.015	nein
Fahrerbeurteilung vorausschauende Fahrweise	0.230	.818	ja	2.011	.048	nein
Fahrerbeurteilung rasante Fahrweise	1.172	.243	ja	-1.312	.194	ja
Fahrerbeurteilung hektische Fahrweise	0.550	.583	ja	-2.689	.009	nein
Beifahrerbeurteilung sichere Fahrweise	-1.097	.275	ja	-1.573	.121	ja
Beifahrerbeurteilung vorausschauende Fahrweise	-0.101	.920	ja	-1.401	.167	ja
Beifahrerbeurteilung rasante Fahrweise	-0.243	.808	ja	-0.484	.630	ja
Beifahrerbeurteilung hektische Fahrweise	-0.016	.987	ja	0.012	.991	ja
<b>Resultateebene</b>						
Kognitive Irritation	3.102	.002	nein	2.762	.007	nein
Emotionale Irritation	1.699	.092	nein	1.970	.053	nein
Irritation gesamt	2.533	.012	nein	2.735	.008	nein
GBB Erschöpfung	0.748	.456	ja	-0.039	.969	ja
GBB Magenbeschwerden	0.650	.517	ja	1.409	.163	ja
GBB Gliederschmerzen	1.871	.063	nein	1.562	.123	ja
GBB Herzbeschwerden	1.130	.260	ja	0.699	.487	ja
GBB Beschwerdedruck	1.184	.239	ja	0.868	.388	ja
Wohlbefinden (WHO-5)	-2.048	.042	nein	-1.868	.066	nein
Burnout (emotionale Erschöpfung)	0.648	.518	ja	-0.116	.908	ja
Burnout (Zynismus)	0.223	.824	ja	-0.864	.390	ja
Vitalität (BBS)	-4.784	.000	nein	-5.110	.000	nein
Gleichgewicht (BBS)	0.286	.775	ja	2.034	.045	nein
Extraversion (BBS)	0.199	.842	ja	0.442	.660	ja
Vigilanz (BBS)	-0.240	.811	ja	1.290	.201	ja
NASA geistige Anforderungen	2.954	.004	nein	3.052	.003	nein
NASA körperliche Anforderungen	2.093	.038	nein	1.342	.184	ja
NASA zeitliche Anforderungen	1.803	.073	ja	0.628	.532	ja
NASA Leistung	0.065	.948	ja	0.958	.341	ja
NASA Anstrengung	2.568	.011	nein	1.835	.071	nein
NASA Frustration	0.011	.991	ja	-0.521	.604	ja
Mittlere Müdigkeit (KSS)	1.075	.284	ja	0.889	.377	ja
Mittlere Ermüdung (KSS)	-0.063	.950	ja	-1.525	.132	ja
Mittlere Herzfrequenz (Schläge pro Minute)	1.085	.280	ja	1.252	.216	ja

Abhängige Variable	Voraussetzung 1: Unterschied T2 Gruppe 2 zu T1 Gruppe 1			Voraussetzung 2: Veränderung über die Zeit T1/T2 in Gruppe 2		
	T-Wert	p-Wert	erfüllt	T-Wert	p-Wert	erfüllt
Maximale Herzfrequenz (Schläge pro Minute)	1.912	.058	nein	1.297	.200	ja
Minimale Herzfrequenz (Schläge pro Minute)	0.968	.335	ja	-0.364	.717	ja
„Hilfsfrist“ (Dauer der Einsatzfahrten, in mm:ss)	0.814	.417	ja	-76.148	.000	nein

\* Fahrprofildaten beziehen sich lediglich auf Hinfahrten zum Einsatzort mit RTW; alle T-Test korrigiert für Varianzheterogenität

In den weiteren Abschnitten werden entsprechend und erneut auf den verschiedenen Evaluationsebenen die (Einzel-)Befunde für die Wartegruppe vorgestellt sowie jeweils Verläufe mit gepaarten und ungepaarten Stichproben.

In Kapitel 4.2 wurden alle Ergebnisse der Hauptanalysen mit Wartekontrollgruppen-Design vorgestellt. Werden die Befunde zusammengefasst, ergibt sich ein gemischtes Bild:

- Alle Teilnehmer bewerten das Training direkt nach der Intervention insgesamt sehr positiv und auch nach längerer Zeit noch recht gut. Die Wartegruppe evaluierte das Training dabei noch besser als die Interventionsgruppen.
- Auf der Lernebene fanden sich durch das Training ein deutlicher Wissenszuwachs bei den Interventionsgruppen im Vergleich zur Wartegruppe sowie eine Zunahme der Risikowahrnehmung bei Normalfahrten.
- Auf der Verhaltensebene sanken die mittleren und maximalen Geschwindigkeiten bei Einsatzfahrten durch das Training. Weitere Fahrprofilparameter änderten sich nicht.
- Auf der Resultateebene ließen sich keine Effekte des Trainings feststellen. Auch die „Hilfsfrist“ veränderte sich nicht.
- Für körperliche Beschwerden und Herzfrequenz zeigten sich in allen Studiengruppen schlechte Werte.
- Auf allen Evaluationsebenen spielten Alter und Geschlecht eine Rolle und wirkten z.T. differenziell auf die Trainingseffekte ein.
- Messeffekte im Vergleich zwischen Gruppe 1 und 3 lassen sich nicht mit Bestimmtheit identifizieren: Bei einer Vielzahl von Testungen finden sich nur wenige Unterschiede, u.a. ist die positive Längsbeschleunigung in der (nicht vorgemessenen) Gruppe 3 deutlicher höher als in Gruppe 1.
- Voraussetzungen für die Vereinfachung des Designs in den folgenden Analysen (durch Zusammenführung der Gruppe 1 und 2) sind nicht gegeben.

### 4.3 Weitere Ergebnisse zur Lernebene

Im folgenden Abschnitt werden die weiteren Ergebnisse zur Lernebene vorgestellt. Zum einen werden die potenziellen Effekte des Trainings in der Wartegruppe analysiert, um festzustellen, ob sie sich ähnlich verhält, wie die Gruppen 1 und 3, d.h. es wird die Replizierbarkeit der Effekte betrachtet. Zum anderen wird die Nachhaltigkeit von Effekten anhand von Vergleichen mit den der Follow-up-Messungen untersucht. Schließlich wird das Lernen durch das Training pro Messzeitpunkt auch mit ungepaarten Stichproben dargestellt, um abzusichern, dass Effekte oder Befunde nicht ausschließlich durch fehlende Werte bedingt

sind. Dieser Aufbau wiederholt sich auch in den darauf folgenden Abschnitten zur Verhaltens- und Resultateebene.

### 4.3.1 Lernen durch das Training in der Kontrollgruppe

Die Analysen in diesem Abschnitt sollen zeigen, ob es für die Gruppe 2 Effekte des Trainings gibt und ob diese die Befunde der Interventionsgruppen 1 und 3 replizieren. Für die Effekte auf das Wissen kann diese Analyse nicht durchgeführt werden, da es nur zweimal erhoben wurde (vgl. Abschnitt 3.2.2). Tabelle 33 zeigt die Daten für den Zeiteffekt in Gruppe 2 kontrolliert für Alter und Geschlecht. Zusätzlich werden die multivariaten Tests der paarweisen Vergleiche angegeben und in den Mittelwertspalten angezeigt, zwischen welchen Zeitpunkten es Unterschiede gibt. Um einen Trainingseffekt anzuzeigen, sollten sich jeweils Zeitpunkt T2 und Zeitpunkt T3 voneinander unterscheiden.

Im Bereich der *verkehrsrelevanten Persönlichkeitseigenschaften* finden sich signifikante Tests der paarweisen Vergleiche für Extraversion und Unverträglichkeit, bei beiden ist allerdings kein eindeutiger Unterschied zwischen T2 und T3 zu erkennen, sondern eher ein „lineares“ Absinken der Werte über alle drei Messzeitpunkte. Darüber hinaus zeigt die verkehrsbezogene Extraversion zeigt einen Geschlechtseinfluss auf die Bewertung der Skala ( $F_{(1,61)}=3.52$ ,  $p=.065$ ). Die Männer zeigen, wie auch schon in den bisherigen Auswertungen, höhere Extraversionswerte und einen Trend zu Abnahme der Werte, während dies bei Frauen nicht der Fall ist. Auch die Unverträglichkeit weist den gleichen Alterseffekt auf, wie er sich in der Hauptauswertung in Abschnitt 4.2.2.1 gezeigt hat ( $F_{(1,61)}=2.84$ ,  $p=.062$ ). Die jüngeren geben insgesamt höhere Werte für Unverträglichkeit an als ihre älteren Kollegen. Insgesamt nimmt die Unverträglichkeit über die Zeit ab. Damit zeigen sich zwar deskriptiv ähnliche Veränderungen in der Wartegruppe wie in den Interventionsgruppen, dennoch lassen sich diese statistisch nicht absichern, so dass eine Replizierbarkeit verneint werden muss.

Tabelle 33: Lernebene: Zeitlicher Verlauf über die ersten drei Messzeitpunkte in der Wartegruppe

Abhängige Variable	N	Zeitpunkt T1 (a)		Zeitpunkt T2 (b)		Zeitpunkt T3 (c)		Kontrollierter Zeiteffekt		Paarweise Vergleiche	
		M	SD	M	SD	M	SD	F	p	F	p
<b>Verkehrsrelevante Persönlichkeitseigenschaften</b>											
Extraversion	64	1.65	0.39	1.62	0.38	1.56	0.43	0.86	.415	<b>2.86</b>	.065 <sup>+</sup>
Sensation Seeking	64	2.36	0.41	2.35	0.39	2.30	0.46	0.92	.401	0.58	.563
Gewissenhaftigkeit	64	2.73	0.59	2.63	0.50	2.69	0.53	1.59	.208	1.51	.230
Unverträglichkeit	64	2.35 <sup>(c)</sup>	0.54	2.28	0.47	2.24 <sup>(a)</sup>	0.48	0.07	.937	<b>2.69</b>	.076 <sup>+</sup>
Reaktanz	64	1.50	0.35	1.51	0.37	1.48	0.34	3.09	.049*	0.32	.728
<b>Risikobewusstsein</b>											
Risikowahrnehmung bei schweren Regelverstößen	64	4.43	0.28	4.43	0.33	4.49	0.30	1.14	.322	<b>2.09</b>	.132
Risikowahrnehmung bei mittelschweren Regelverstößen	64	3.57 <sup>c</sup>	0.55	3.64	0.61	3.73 <sup>a</sup>	0.57	1.85	.162	<b>4.72</b>	.012*
Risikowahrnehmung bei leichten Regelverstößen	64	3.27 <sup>c</sup>	0.51	3.25 <sup>c</sup>	0.58	3.41 <sup>a,b</sup>	0.52	0.13	.876	<b>7.41</b>	.001***
Risikowahrnehmung bei normalem Fahren	64	2.68 <sup>c</sup>	0.58	2.68 <sup>c</sup>	0.54	2.84 <sup>a,b</sup>	0.55	1.37	.257	<b>6.03</b>	.004**
Risk Seeking	64	1.81 <sup>b,(c)</sup>	0.70	1.62 <sup>a</sup>	0.58	1.65 <sup>(a)</sup>	0.62	1.51	.227	<b>3.79</b>	.028*
<b>Subjektive Fahrfähigkeit</b>											
Fahrtkompetenz	64	7.52	1.15	7.34	1.12	7.31	0.97	0.38	.683	1.72	.188

Abhängige Variable	N	Zeitpunkt T1 (a)		Zeitpunkt T2 (b)		Zeitpunkt T3 (c)		Kontrollierter Zeiteffekt		Paarweise Vergleiche	
		M	SD	M	SD	M	SD	F	p	F	p
allgemein											
Fahrtkompetenz Rettungsdienst	64	7.03	1.40	6.80	1.39	6.92	1.24	1.53	.224	1.25	.295
Unfallwahrscheinlichkeit allgemein	64	4.88	2.22	5.28	1.91	5.06	1.99	1.87	.158	1.12	.333
Unfallwahrscheinlichkeit Rettungsdienst	64	4.80	2.00	5.25	2.05	4.75	1.88	<b>3.96</b>	.027*	1.39	.257

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Kontraste zur Berechnung der Zeiteffekte: einfach (T3 zu T1 und T3 zu T2) und umgekehrter Helmert (T3 zu Mittelwert aus T1 und T2); Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Zeitpunkten

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η² ≥ .06)

Für die Gruppe von Variablen zum *Risikobewusstsein* zeigten sich im Hauptdesign erwünschte Trainingseffekte auf die Risikowahrnehmung bei Normalfahrten sowie tendenziell auch auf die Risikowahrnehmung bei leichten Verstößen. Tabelle 33 zeigt in der Wartegruppe für alle Aspekte dieser Variablengruppe – mit Ausnahme der Risikowahrnehmung bei schweren Verstößen - signifikante paarweise Vergleiche, die bei den schon im Hauptdesign bedeutsamen Variablen hier auch zwischen T2 und T3 liegen, d.h. die Erhöhung der Risikowahrnehmung bei Normalfahrten und leichten Verstößen lässt sich als Trainingseffekt replizieren. Risk Seeking ist in der Wartegruppe zu T1 am höchsten ausgeprägt und nimmt dann bereits vor dem Training ab und scheint nach dem Training stabil zu bleiben – hier kann also nicht von einem Trainingseffekt ausgegangen werden.

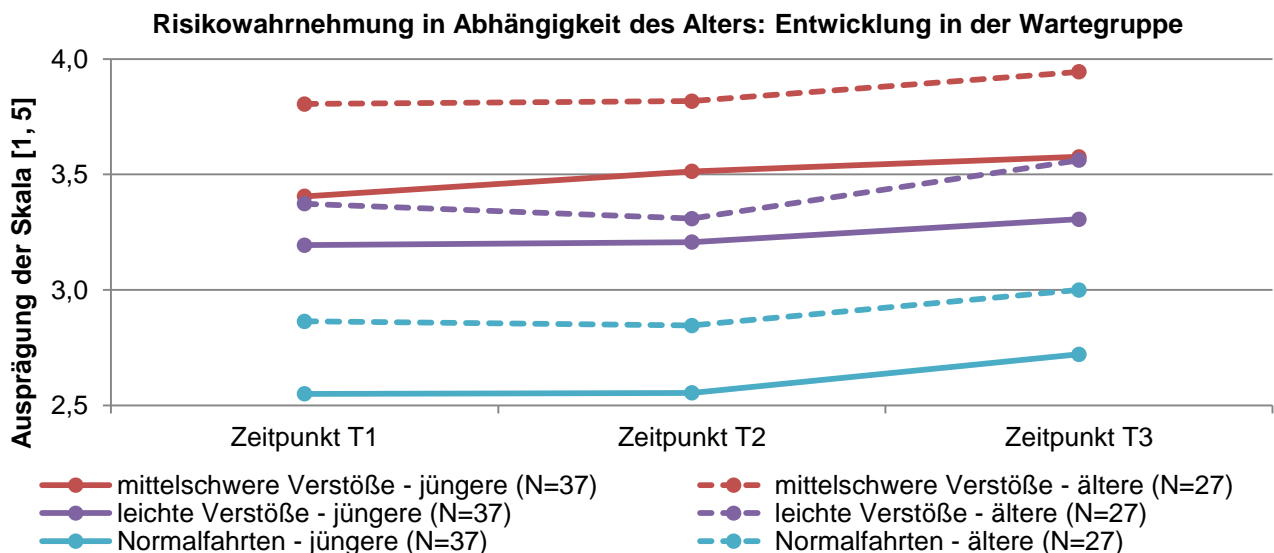


Abbildung 40: Risikowahrnehmung für mittelschwere und leichte Verstöße sowie bei Normalfahrten in der Wartegruppe in Abhängigkeit vom Alter

Fast alle Aspekte des Risikobewusstseins weisen, wie auch schon bei den Hauptauswertungen gefunden, signifikante oder tendenzielle Alterseffekte auf (mittelschwere Verstöße:  $F_{(1,61)}=12.00$ ,  $p=.001$ , leichte Verstöße:  $F_{(1,61)}=6.91$ ,  $p=.011$ , Normalfahrten: Alter  $F_{(1,61)}=11.3$ ,  $p=.001$ ). Abbildung 40 zeigt, dass die älteren Probanden der Wartegruppe insgesamt ein höheres Risikobewusstsein angeben und mit der Schwere der

Verstöße das wahrgenommene Risiko steigt. Zwischen den Zeitpunkten T2 und T3 ist der schon dargestellte Trend in die erwünschte Richtung zu sehen.

Für den Bereich der subjektiven Fahrfähigkeit fanden sich in den Hauptanalysen keine eindeutigen Trainingseffekte. Dies ist auch für die Wartegruppe der Fall, es zeigen sich keine signifikanten Änderungen zwischen T2 und T3. Der einzige Effekt, der sich in den erfragten subjektiven Fahrfähigkeiten finden lässt, ist ein Zeiteffekt für die wahrgenommene eigene Unfallwahrscheinlichkeit als Fahrer im Rettungsdienst. Im zeitlichen Verlauf steigt diese in der Zeit zwischen T1 und T2 an und sinkt nach dem Training wieder ab.

In der Gesamtschau sind damit vor allem die Veränderungen im Risikobewusstsein durch das Training replizierbar und belastbar, aber auch die im Hauptdesign gefundenen Effekte der Kontrollvariablen Alter und Geschlecht scheinen sich in der Wartegruppe zu wiederholen.

### 4.3.2 Lernen im langfristigen Verlauf

Um die Nachhaltigkeit von Effekten oder Effekte, die erst mit längerer Verzögerung auftreten zu analysieren, werden die Ergebnisse der Follow-up-Messungen (T4) im Unterschied zur Ersterhebung (T1) über alle Gruppen hinweg betrachtet. Die einfaktoriellen Varianzanalysen wurden erneut für Alter und Geschlecht sowie für die Zugehörigkeit zu den Studiengruppen kontrolliert. Tabelle 34 gibt einen Überblick über diese Analysen für die Variablen der Lernebene.

Tabelle 34: Lernebene: Langfristiger Verlauf in der Gesamtstichprobe

Abhängige Variable über alle Gruppen	N	Zeitpunkt T1 (a)		Zeitpunkt T4 (b)		Zeiteffekt	
		M	SD	M	SD	F	p
Extraversion	45	1.61	0.44	1.60	0.46	0.43	.514
Sensation Seeking	45	2.36	0.43	2.27	0.45	0.87	.355
Gewissenhaftigkeit	44	2.67	0.59	2.67	0.57	0.14	.707
Unverträglichkeit	45	2.42	0.49	2.32	0.53	<b>3.40</b>	.072 <sup>+</sup>
Reaktanz	45	1.49	0.50	1.52	0.44	2.01	.164
Risikowahrnehmung bei schweren Regelverstößen	45	4.46	0.35	4.52	0.30	0.58	.449
Risikowahrnehmung bei mittelschweren Regelverstößen(*)	45	3.73	0.55	3.85	0.48	0.18	.674
Risikowahrnehmung bei leichten Regelverstößen	45	3.46	0.60	3.51	0.45	0.19	.669
Risikowahrnehmung bei normalem Fahren	45	2.78	0.52	2.84	0.50	0.04	.847
Risk Seeking	45	1.97	0.86	1.82	0.81	0.10	.751
Fahrtkompetenz allgemein	45	7.22	1.20	7.23	1.30	0.05	.819
Fahrtkompetenz Rettungsdienst(*)	45	6.95	1.37	7.26	1.50	0.00	.998
Unfallwahrscheinlichkeit allgemein	45	4.51	2.01	4.64	1.82	0.03	.857
Unfallwahrscheinlichkeit Rettungsdienst	45	5.02	1.96	4.56	1.84	0.00	.946

Berechnung: Varianzanalyse (Kontrolle von Alter, Geschlecht und Versuchsgruppe); Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; (\*) hinter den Skalen kennzeichnet Unterschiede zwischen T1 und T4, die bei unkontrollierter Analyse mit 10% Irrtumswahrscheinlichkeit signifikant sind

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Bei den *verkehrsrelevanten Persönlichkeitseigenschaften* zeigt sich nur für die Unverträglichkeit eine leichte längerfristige Abnahme der Werte ( $\eta^2_{\text{part}}=.077$ ). Für Extraversion und Reaktanz zeigen sich keinerlei länger-

fristige Effekte des Trainings, bei Sensation Seeking, Gewissenhaftigkeit und Unverträglichkeit zeigen sich erneut einige Einflüsse der Kontrollvariablen, die aber aufgrund der sehr kleinen Gruppengrößen nicht weiter betrachtet, zumal sie nicht auf das Training zurückgeführt werden können.

Hinsichtlich des *Risikobewusstseins* gibt es insgesamt keine Unterschiede zwischen der ersten Messung und der Follow-up-Messung, wenn Alter, Geschlecht und die Gruppenzuordnung kontrolliert werden. Deskriptiv liegen zwar alle Werte der Skalen zur Risikowahrnehmung zu T4 höher als zu T1, aber die Unterschiede werden unter Kontrolle nicht signifikant, d.h. auch hier spielen die Kontrollvariablen eine deutliche Rolle, die beschreibend vergleichbar den bisherigen Analysen ist (u.a. höhere Risikowahrnehmung bei älteren Probanden), aber aufgrund der kleinen Gruppengrößen inferenzstatistisch nicht mehr zuverlässig abbildbar ist. Damit zeigt sich zwar nach wie vor ein positiver Trend für eine Entwicklung des Risikobewusstseins in die gewünschte Richtung, der signifikante kurzfristige Effekt des Trainings auf die Risikowahrnehmung bei Normalfahrten und bei leichten Verstößen schwächt sich jedoch über die Monate nach dem Training wieder ab.

Im Hinblick auf die subjektive Fahrfähigkeit zeigte sich im Hauptdesign ein signifikanter Trainingseffekt für die allgemeine Fahrkompetenz, dieser Effekt ist nicht nachhaltig – weder für diese Einschätzung noch für die weiteren Einschätzungen der eigenen Fahrfähigkeit zeigen sich relevante längerfristige Veränderungen.

### 4.3.3 Lernen durch das Training in der ungepaarten Stichprobe

Um potenzielle Verzerrungen der Daten aufgrund der vielen Dropouts zu unterschiedlichen Zeitpunkten und in verschiedenen Variablen zu kontrollieren, werden im Folgenden Analysen mit den Messzeitpunkten als unabhängigem Faktor (ungepaarte Stichprobe) betrachtet. So können alle Probanden, die einzelne Werte haben, in die Berechnungen aufgenommen werden. Da bereits einige Berechnungen durchgeführt wurden, soll dieser Abschnitt nur dazu dienen, zu überprüfen, ob die Effekte, die bereits gefunden wurden auch bestehen bleiben, wenn die Dropouts mit betrachtet werden. Zusätzlich soll überprüft werden, ob Effekte, die bisher nicht entdeckt werden konnten, hinzukommen. Daher werden im folgenden Abschnitt nur die Variablen aufgeführt, bei denen die Effekte relevant sind und nicht mehr alle abhängigen Variablen.

Tabelle 35: Lernebene: Kurzfristige Effekte in der ungepaarten Stichprobe

Abhängige Variable	N <sub>T1</sub> /N <sub>T2</sub>	Gruppeneffekt		Effekt des Zeitpunkts		Interaktionseffekt	
		F	p	F	p	F	p
Wissen	177/170	0.54	.584	5.40	.021*	6.16	.002**
Extraversion	179/165	0.20	.816	0.00	.975	1.34	.263
Risikowahrnehmung bei mittelschweren Regelverstößen	179/165	4.67	.010**	3.27	.072 <sup>+</sup>	0.13	.876
Risikowahrnehmung bei leichten Regelverstößen	179/165	5.68	.004**	1.57	.211	0.55	.580
Risikowahrnehmung bei normalem Fahren	179/165	3.60	.028*	3.54	.061 <sup>+</sup>	1.30	.273
Fahrtkompetenz allgemein	178/165	4.45	.012*	0.06	.806	0.63	.534

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht) mithilfe der ungepaarten Stichprobe über die Zeitpunkte T1 und T2 zwischen den Gruppen und Zeitpunkten; Abkürzungen: N=Anzahl der Probanden zu den zwei Zeitpunkten; F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Zunächst werden die kurzfristigen Haupteffekte verglichen, also die Unterschiede zwischen den Gruppen zu den Zeitpunkten T1 und T2. Ein Trainingseffekt konnte nur bei dem Wissenstest und bei der Risikowahrnehmung bei normalen Fahrten gefunden werden. Einen Interaktionseffekt, der jedoch nicht eindeutig auf das Training schließen ließ, gab es bei der Extraversion und der allgemeinen Fahrkompetenz. Die Berechnungen mit der Gruppe 2 im langfristigen Verlauf ließen zusätzlich Trainingseffekte auf die Risikowahrnehmung bei leichten Verstößen vermuten. Alle anderen Effekte waren unabhängig vom Training. Neben diesen fünf Variablen werden in Tabelle 35 zusätzlich alle Variablen berichtet, die „neue“ Effekte unter Einschluss der Dropout zeigen.

Der Trainingseffekt für den Wissenstest kann auch mit der ungepaarten Stichprobe bestätigt werden: Die Hinzunahme sieben weiterer Vorherstestungen führt zum gleichen Ergebnis, die Interventionsgruppen 1 und 3 haben durch das Training einen Wissenszuwachs, wohingegen die Wartegruppe in ihren Werten leicht sinkt.

Der Effekt des Trainings auf die Extraversion, der nur in der Wartegruppe gefunden wurde (T2-T3-Vergleich), verschwindet, wenn alle vorliegenden Daten berücksichtigt werden. Auch die Interaktion, die in den Haupteffekten aufgrund der Änderungen in Gruppe 3 gefunden wurden, wird nicht mehr signifikant durch Hinzunahme der Dropouts. Die Risikowahrnehmung bei leichten Verstößen stieg nach dem Training nur in der Wartegruppe an und zeigte ansonsten einen nicht trainingsbezogenen Gruppenunterschied. Dieser bleibt auch bei der Analyse mit vollständigen Daten bestehen. Der Trainingseffekt auf die Risikowahrnehmung bei Normalfahrten drückt sich (vermutlich aufgrund des linearen Verhaltens der beiden Interventionsgruppen) nicht mehr in einer Interaktion aus, dennoch zeigen Zeitpunkt- und Gruppenhaupteffekte denselben Verlauf wie bei den Hauptanalysen mit gepaarter Stichprobe (vgl. Abbildung 41). Neu hinzugekommen und mit vergleichbarer Ausrichtung zeigt sich im Bereich des Risikobewusstseins ein Gruppen- und Zeitpunkteffekt für die Risikowahrnehmung bei mittelschweren Verstößen. Für die Einschätzung der allgemeinen eigenen Fahrkompetenz fand sich im Längsschnitt ein Interaktionseffekt, der einen Zuwachs ausschließlich in Gruppe 1 zeigte. Dieser Effekt zeigt sich unter Hinzunahme der vollständigen Daten nicht mehr. Generell bleiben unter Hinzunahme der Dropouts die „belastbarsten“ Befunde zu kurzfristigen Effekten auf der Lernebene bestehen, bereits in der längsschnittlichen Analyse uneindeutigere Befunde zeigen sich hingegen volatiler.

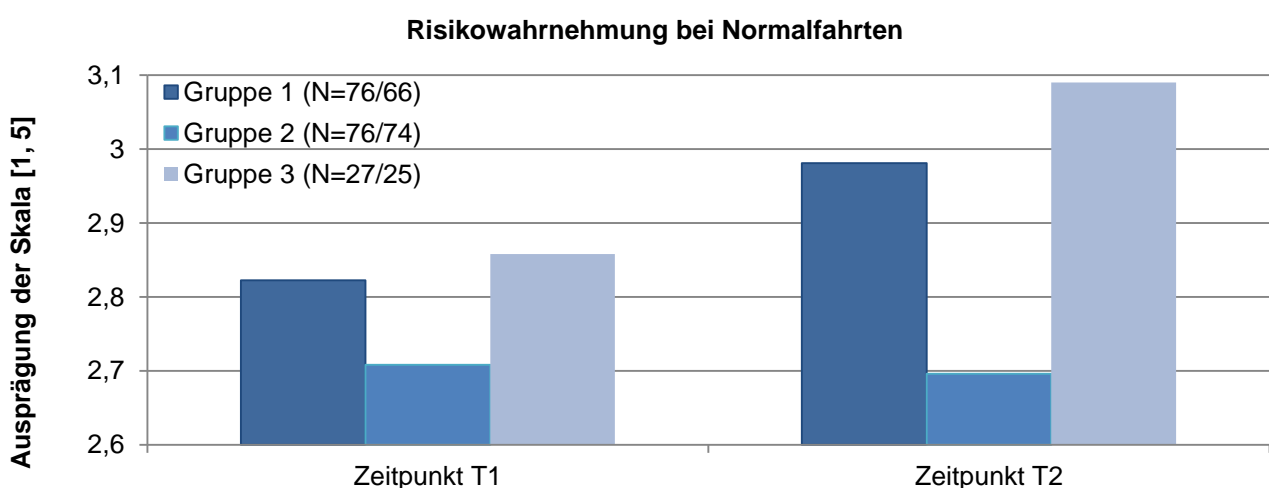


Abbildung 41: Risikowahrnehmung bei Normalfahrten in der ungepaarten Stichprobe

Im langfristigen Verlauf konnten in den gepaarten Hauptanalysen nur Trainingseffekte auf die Unverträglichkeit gefunden werden, ohne dass dieser Effekt in den T1-T2 Vergleichen zu sehen gewesen wäre, so



dass unklar ist, ob die Abnahme ursächlich auf das Training zurückzuführen ist. Dennoch wird diese Variable unter Hinzunahme der Dropouts noch einmal betrachtet. Zusätzlich werden in Tabelle 36 alle Variablen aufgenommen, in denen sich weitere Haupteffekte ändern oder die langfristig im ungepaarten Vergleich zusätzliche Effekte zeigen.

Der Effekt auf die Unverträglichkeit verschwindet, wenn die Dropouts in die Berechnungen einbezogen werden. Zusätzlich verschwinden auch die Gruppenunterschiede bei Sensation Seeking und der Risikowahrnehmung bei schweren Verstößen.

Tabelle 36: Lernebene: Langfristige Effekte in der ungepaarten Stichprobe

Abhängige Variable	N <sub>T1</sub> /N <sub>T4</sub>	Gruppeneffekt		Effekt des Zeitpunkts		Interaktionseffekt	
		F	p	F	p	F	p
Sensation Seeking	179/45	1.89	.154	0.01	.924	2.44	.090 <sup>+</sup>
Gewissenhaftigkeit	178/45	3.27	.040*	1.48	.225	0.90	.409
Unverträglichkeit	179/45	0.20	.820	0.06	.813	0.15	.864
Risikowahrnehmung bei schweren Regelverstößen	179/45	0.71	.495	0.35	.553	0.81	.447
Risk Seeking	178/45	3.31	.038*	0.00	.953	0.83	.439
Fahrtkompetenz allgemein	178/45	4.57	.011*	0.13	.723	0.14	.873
Fahrtkompetenz Rettungsdienst	178/45	3.41	.035*	2.65	.105	0.21	.814

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht) mithilfe der gepaarten Stichprobe über die Zeitpunkte T1 und T2 zwischen den Gruppen und Zeitpunkten; Abkürzungen: N=Anzahl der Probanden zu den zwei Zeitpunkten; F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

Statt einem Gruppenunterschied zeigen die ungepaarten Analysen nun einen Interaktionseffekt zwischen Gruppe und Zeitpunkt in die Richtung, dass das Sensation Seeking in Gruppe 1 abnimmt während es in Gruppe 2 und 3 zwischen T1 und T4 steigt. Bei der Gewissenhaftigkeit, dem Risk Seeking und der wahrgenommenen allgemeinen Fahrtkompetenz bleiben die Gruppenunterschiede bestehen.

Für die Effekte des Trainings auf der Lernebene fanden sich zusammenfassend folgende Befunde:

- Die Befunde zur Erhöhung der Risikowahrnehmung bei Normalfahrten konnten in der Wartegruppe repliziert werden und zeigten sich in dieser Gruppe auch für weitere Aspekte der Risikowahrnehmung.
- Im langfristigen Vergleich und bei der Betrachtung unverbundener Stichproben (d.h. unter Hinzunahme der Dropout-Daten) war die Erhöhung der Risikowahrnehmung nicht mehr signifikant, allerdings blieben die Unterschiede in ihren Trends bestehen.
- Für alle Variablen der Risikowahrnehmung zeigte sich ein klarer Alterseffekt – ältere Teilnehmer bewerteten Risiken im Verkehr durchgängig höher als jüngere.
- Als zusätzlicher Effekt, der nicht in den Hauptanalysen zu finden war, sanken in der Wartegruppe die Unverträglichkeit sowie die Extraversion als verkehrsrelevante Persönlichkeitsmerkmale über die Zeit ab. Der Effekt auf die Unverträglichkeit blieb langfristig bestehen, während dies für die Extraversion nicht der Fall war. Beide Effekte verschwanden auch unter Hinzunahme der Dropout-Daten.

## 4.4 Weitere Ergebnisse zur Verhaltensebene

### 4.4.1 Verhaltensänderung durch das Training in der Kontrollgruppe

Im Folgenden wird, analog des Abschnitts 4.3.1, die Gruppe 2 hinsichtlich der drei gemessenen Zeitpunkte betrachtet. Es wird untersucht, ob sich die Wartegruppe zwischen den Zeitpunkten in den Verhaltensparametern ändert, v.a. ob sich die Hauptergebnisse der Interventionsgruppen im Vergleich von T2 und T3 replizieren lassen. Tabelle 37 zeigt die Ergebnisse der entsprechenden Analysen.

Tabelle 37: Verhaltensebene: Zeitlicher Verlauf über die ersten drei Messzeitpunkte in der Wartegruppe

Abhängige Variable	N	Zeitpunkt T1 (a)		Zeitpunkt T2 (b)		Zeitpunkt T3 (c)		Kontrollierter Zeiteffekt		Paarweise Vergleiche	
		M	SD	M	SD	M	SD	F	p	F	p
		<b>Fahrprofildaten: Geschwindigkeiten</b>									
Durchschnittliche Geschwindigkeit (in km/h)	47	51.49	10.05	53.08	10.04	52.06	10.34	0.67	.495	1.05	.360
Maximale Geschwindigkeit (in km/h)	47	117.10	18.40	118.67	19.64	113.85	21.84	0.37	.657	<b>1.44</b>	.248
<b>Fahrprofildaten: Längsbeschleunigungen</b>											
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	47	3.36	1.92	2.81	1.04	2.92	1.08	2.03	.153	<b>1.46</b>	.244
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	47	0.33 <sup>b</sup>	0.16	0.27 <sup>a</sup>	0.04	0.28	0.06	<b>3.28</b>	.069 <sup>+</sup>	<b>3.23</b>	.049*
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	47	0.06	0.26	0.00	0.01	0.00	0.00	1.62	.210	1.22	.306
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	47	0.03	0.14	0.00	0.01	0.00	0.00	1.84	.181	1.38	.264
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	47	-3.76	1.97	-3.13	1.06	-3.09	0.81	<b>3.32</b>	.057 <sup>+</sup>	2.21	.122
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	47	-0.34 <sup>b</sup>	0.15	-0.29 <sup>a</sup>	0.05	-0.29	0.06	2.69	.099 <sup>+</sup>	<b>3.29</b>	.047*
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	47	1.57	4.98	0.07	0.12	0.11	0.19	<b>3.22</b>	.079 <sup>+</sup>	<b>2.86</b>	.068 <sup>+</sup>
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	47	0.59	1.95	0.01	0.02	0.02	0.04	<b>3.23</b>	.079 <sup>+</sup>	<b>3.89</b>	.028*
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	47	0.06	0.26	0.00	0.01	0.00	0.00	1.43	.238	<b>2.25</b>	.117
Anzahl Trigger pro Minute negative Längsbeschleunigung (5,9 m/s <sup>2</sup> )	47	0.03 <sup>(c)</sup>	0.13	0.00	0.00	0.00 <sup>(a)</sup>	0.00	1.76	.192	<b>2.42</b>	.101
<b>Fahrprofildaten: Querschleunigungen</b>											
Maximale positive Querschleunigung (in m/s <sup>2</sup> )	47	5.37 <sup>b,c</sup>	4.04	3.87 <sup>a</sup>	1.17	3.84 <sup>a</sup>	0.96	1.36	.256	<b>3.31</b>	.046*

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Zeitpunkt T1 (a)		Zeitpunkt T2 (b)		Zeitpunkt T3 (c)		Kontrollierter Zeiteffekt		Paarweise Vergleiche	
		M	SD	M	SD	M	SD	F	p	F	p
Mittlere positive Querbeschleunigung (in m/s <sup>2</sup> )	47	0.50 <sup>b,c</sup>	0.30	0.39 <sup>a</sup>	0.12	0.36 <sup>a</sup>	0.09	1.01	.337	<b>4.81</b>	.013*
Maximale negative Querbeschleunigung (in m/s <sup>2</sup> )	47	-5.05 <sup>c</sup>	4.01	-3.85 <sup>(c)</sup>	1.14	-3.48 <sup>a,(b)</sup>	0.77	1.99	.164	<b>5.66</b>	.007**
Mittlere negative Querbeschleunigung (in m/s <sup>2</sup> )	47	-0.47 <sup>(b),c</sup>	0.30	-0.39 <sup>(a),(c)</sup>	0.11	-0.34 <sup>a,(b)</sup>	0.07	0.96	.349	<b>4.12</b>	.023*
Dauer Trigger Querbeschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	47	1.48	4.88	0.01	0.03	0.00	0.00	2.29	.137	<b>3.29</b>	.047*
Anzahl Trigger pro Minute Querbeschleunigung (6,9 m/s <sup>2</sup> )	47	0.62 <sup>(b),c</sup>	1.98	0.00 <sup>(a)</sup>	0.01	0.00 <sup>(a)</sup>	0.00	2.53	.119	<b>4.10</b>	.024*
Selbsteinschätzung der Einsatzfahrten											
Fahrerbeurteilung sichere Fahrweise	65	82.06	10.61	79.42	14.02	81.35	14.42	0.21	.807	1.61	.209
Fahrerbeurteilung vorausschauende Fahrweise	66	82.56	10.72	81.48	12.81	82.08	11.99	0.08	.925	0.53	.592
Fahrerbeurteilung rasante Fahrweise	66	36.87	19.04	38.47	21.76	36.21	21.30	0.46	.633	0.50	.610
Fahrerbeurteilung hektische Fahrweise	66	14.26	11.65	15.92	12.81	15.84	13.44	0.06	.944	1.34	.270
Fremdbeurteilung der Einsatzfahrten											
Beifahrerbeurteilung sichere Fahrweise	49	87.17	14.25	90.22 <sup>c</sup>	7.39	87.13 <sup>b</sup>	11.32	0.15	.814	<b>2.86</b>	.068 <sup>+</sup>
Beifahrerbeurteilung vorausschauende Fahrweise	49	84.69	12.57	87.51	8.78	85.92	11.61	1.87	.166	1.19	.315
Beifahrerbeurteilung rasante Fahrweise	49	33.01	19.76	36.28	22.34	32.79	22.01	0.07	.934	0.58	.562
Beifahrerbeurteilung hektische Fahrweise	49	14.54	15.16	13.71	11.63	14.64	15.60	0.22	.805	0.08	.926

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Kontraste zur Berechnung der Zeiteffekte: einfach (T3 zu T1 und T3 zu T2) und umgekehrter Helmert (T3 zu Mittelwert aus T1 und T2); Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Zeitpunkten

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)  
Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

Im Bereich der Fahrprofile konnten für die *Geschwindigkeit* im Hauptdesign Trainingseffekte in die Richtung gefunden werden, dass sich die maximale und mittlere Geschwindigkeit der Gruppen 1 und 3 durch das Training verringerten, wohingegen die Gruppe 2 einen leichten Anstieg zeigte. Diese Effekte lassen sich in den Analysen nicht replizieren, dennoch zeigen die Richtung der Entwicklung zwischen T2 und T3 ebenso wie die Stärke der Abnahme einen vergleichbaren Trend wie in den Interventionsgruppen (vgl. Abbildung 42). Zusätzlich gibt es für die maximale Geschwindigkeit einen Geschlechtseffekt dahingehen, dass Männer insgesamt höhere maximale Geschwindigkeiten fahren als Frauen (F<sub>(1,44)</sub>=5.82, p=.034).

Die *Beschleunigungswerte* als weiterer Parameter zeigten in den Hauptanalysen einen Unterschied zwischen den Messzeitpunkten T1 und T2, wenn nicht für Alter und Geschlecht kontrolliert wird. Den größeren Einfluss auf die Längsbeschleunigungswerte hatte die Kontrollvariable Geschlecht dahingehend, dass Männer zu T1 höhere Beschleunigungen auswiesen und Männer und Frauen sich zu T2 annäherten. In den Analysen der Wartegruppe fanden sich für alle Brems-Werte tendenzielle Effekte über die Zeit unter Kontrolle von Alter und Geschlecht. Auch die mittlere positive Beschleunigung zeigt einen tendenziellen Effekt, allerdings zeigen die paarweisen Vergleiche, dass diese Unterschiede die Zeitpunkte T1 und T2 betreffen und damit keinen Effekt des Trainings darstellen.

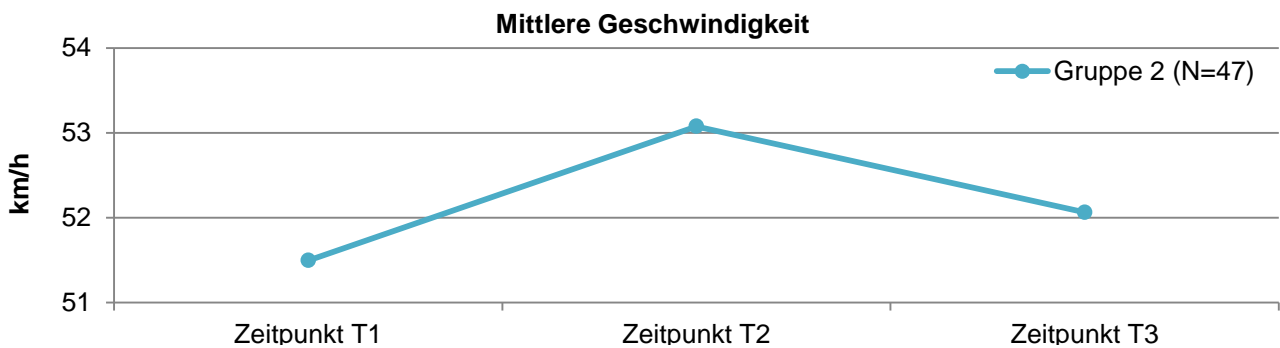


Abbildung 42: Mittlere Geschwindigkeiten bei RTW-Fahrten zum Einsatzort im Zeitverlauf in der Wartegruppe

Die Quereschleunigung in der Wartegruppe zeigt, wie auch die Längsbeschleunigung, Unterschiede zwischen den Zeitpunkten T1 und T2, also bereits vor dem Training. Die kontrollierten Zeiteffekte werden nicht signifikant, allerdings zeigen sich einige signifikante Einzelvergleiche. Eine Verringerung zwischen den Zeitpunkten T2 und T3 findet sich bei der negativen Quereschleunigung (Rechtskurven). Dieser Parameter sank allerdings auch schon ohne Trainingsbeteiligung zwischen T1 und T2.

Im Bereich der Fahrprofile bleiben damit relativ belastbar als Trainingseffekte Verringerungen in den Geschwindigkeiten bestehen, während alle anderen Parameter keine eindeutig auf das Training zurückführbare Effekte zeigen.

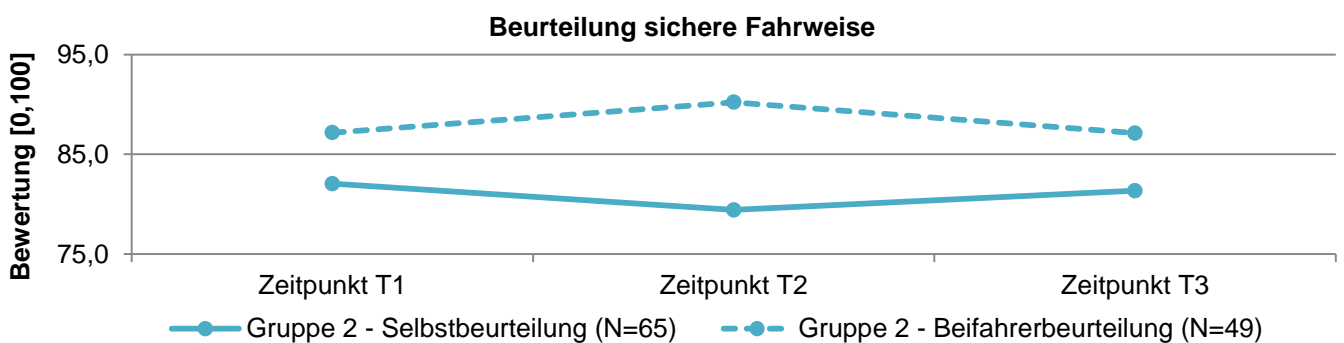


Abbildung 43: Selbst- und Fremdbeurteilung der sicheren Fahrweise im zeitlichen Verlauf für die Wartegruppe

Für die *Selbst- und Fremdeinschätzung der Einsatzfahrten* fanden sich im Hauptdesign keinerlei eindeutigen Trainingseffekte - die Beifahrer beurteilten die Fahrer durchgängig besser als sie sich selbst und die jüngeren Fahrer wurden zum Zeitpunkt T2 besser bewertet als zum Zeitpunkt T1. Im zeitlichen Verlauf der Gruppe 2 findet sich lediglich für die Beifahrereinschätzung der sicheren Fahrweise ein Unterschied zwischen den Zeitpunkten T2 und T3. Die Bewertung sinkt auf das gleiche Niveau, auf dem sie zum

Zeitpunkt T1 lag. Interessanterweise ist der Zeitpunkt T2 in der Selbsteinschätzung geringer als an den anderen Zeitpunkten (allerdings ohne jegliche statistische Bedeutsamkeit), in der Beifahrerbeurteilung aber am höchsten (vgl. Abbildung 43). Insgesamt verhalten sich die Einschätzungen der Fahrweise vor und nach dem Training in der Wartegruppe damit völlig vergleichbar zu den Einschätzungen in den Interventionsgruppen – ein Effekt des Trainings auf diese Variablen ist nicht zu finden.

#### 4.4.2 Verhaltensänderung im langfristigen Verlauf

Der langfristige Verlauf des Verhaltens wird erneut mithilfe der Follow-up-Messungen (T4) im Vergleich zu den ersten Messungen (T1) überprüft. Die Tabelle 34 beschreibt die Mittelwerte zu den Messzeitpunkten über alle Gruppen sowie die Veränderungs-/Zeiteffekte.

Tabelle 38: Verhaltensebene: Langfristiger Verlauf in der Gesamtstichprobe

Abhängige Variable über alle Gruppen	N	Zeitpunkt T1 (a)		Zeitpunkt T4 (b)		Zeiteffekt	
		M	SD	M	SD	F	p
Durchschnittliche Geschwindigkeit (in km/h)	29	53.49	9.77	50.85	8.06	<b>4.53</b>	.043*
Maximale Geschwindigkeit (in km/h)	29	118.00	17.50	113.52	21.60	<b>10.99</b>	.003**
Maximale positive Längs-beschleunigung (in m/s <sup>2</sup> ) (*)	29	4.09	2.70	3.17	1.06	1.43	.242
Mittlere positive Längs-beschleunigung (in m/s <sup>2</sup> ) (*)	29	0.39	0.16	0.29	0.04	<b>3.14</b>	.089 <sup>+</sup>
Dauer Trigger positive Längs-beschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	29	0.07	0.23	0.00	0.01	1.00	.326
Anzahl Trigger pro Minute positive Längs-beschleunigung (5,9 m/s <sup>2</sup> )	29	0.03	0.11	0.00	0.00	0.90	.353
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> ) (*)	29	-4.06	2.21	-3.44	1.21	<b>5.03</b>	.034*
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> ) (*)	29	-0.39	0.15	-0.29	0.05	0.88	.356
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> ) (*)	29	1.92	4.21	0.32	0.57	<b>2.75</b>	.110
Anzahl Trigger pro Minute negative Längs-beschleunigung (3 m/s <sup>2</sup> ) (*)	29	0.70	1.60	0.04	0.05	<b>2.68</b>	.114
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	29	0.07	0.28	0.01	0.03	1.24	.276
Anzahl Trigger pro Minute negative Längs-beschleunigung (5,9 m/s <sup>2</sup> )	29	0.03	0.13	0.00	0.01	1.37	.253
Maximale positive Quer-beschleunigung (in m/s <sup>2</sup> ) (*)	29	6.52	4.72	3.98	1.52	0.72	.403
Mittlere positive Quer-beschleunigung (in m/s <sup>2</sup> ) (*)	29	0.63	0.35	0.38	0.12	0.01	.912
Maximale negative Quer-beschleunigung (in m/s <sup>2</sup> ) (*)	29	-6.14	4.81	-3.92	1.45	0.54	.468
Mittlere negative Quer-beschleunigung (in m/s <sup>2</sup> ) (*)	29	-0.59	0.35	-0.34	0.06	0.02	.887
Dauer Trigger Quer-beschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> ) (*)	29	2.00	5.14	0.05	0.29	<b>2.33</b>	.139

Abhängige Variable über alle Gruppen	N	Zeitpunkt T1 (a)		Zeitpunkt T4 (b)		Zeiteffekt	
		M	SD	M	SD	F	p
Anzahl Trigger pro Minute Querb beschleunigung (6,9 m/s <sup>2</sup> ) (*)	29	0.87	2.12	0.00	0.02	<b>2.59</b>	.120
Fahrerbeurteilung sichere Fahrweise	34	81.94	10.65	82.41	11.65	0.01	.912
Fahrerbeurteilung voraus-schauende Fahrweise	36	82.65	9.95	83.54	9.55	0.01	.927
Fahrerbeurteilung rasante Fahrweise	36	42.44	22.68	38.43	27.35	0.71	.405
Fahrerbeurteilung hektische Fahrweise	36	16.01	14.42	18.29	16.92	0.54	.468
Beifahrerbeurteilung sichere Fahrweise	30	88.10	9.66	88.08	9.35	0.37	.548
Beifahrerbeurteilung vor-ausschauende Fahrweise	30	85.03	13.76	88.19	9.71	1.15	.293
Beifahrerbeurteilung rasante Fahrweise	30	32.86	21.07	32.39	20.71	0.22	.642
Beifahrerbeurteilung hektische Fahrweise	30	16.72	15.94	13.06	13.57	0.71	.406

Berechnung: Varianzanalyse (Kontrolle von Alter, Geschlecht und Versuchsgruppe); Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; (\*) hinter den Skalen kennzeichnet Unterschiede zwischen T1 und T4, die bei unkontrollierter Analyse mit 10% Irrtumswahrscheinlichkeit signifikant sind

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ ) Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

Im Bereich der maximalen *positiven Längsbeschleunigung* zeigt sich unkontrolliert eine Abnahme der Werte, während dieser Effekt kontrolliert für Alter, Geschlecht und Gruppe verschwindet. Bei der mittleren positiven Längsbeschleunigung zeigen sich sowohl für das kontrollierte wie für das unkontrollierte Modell längerfristige Abnahmen, die so in der kurzfristigen Evaluation nicht zu sehen waren. Wie oft Probanden in der positiven Längsbeschleunigung in kritische Bereiche kommen, unterscheidet sich zwischen der ersten Erhebung und der letzten Erhebung nicht.

Für die maximale *negative Längsbeschleunigung* (Bremsen) zeigt sich langfristig ebenfalls eine Abnahme der Werte, die in den vorherigen Analysen nicht zu sehen war, und die Brems-Trigger entwickeln sich entsprechend der maximalen Bremsverzögerung.

Für die *Querb beschleunigung* finden sich kontrolliert für Alter, Geschlecht und Gruppe keine signifikanten Veränderungen zwischen den Zeitpunkten T1 und T4, allerdings sind in den unkontrollierten Analysen Unterschiede dahingehend zu finden, dass die Querb beschleunigung über alle verwendeten Parameter hinweg zum Zeitpunkt T4 geringer als zum Zeitpunkt T1 ist. Die Trigger der Querb beschleunigung, sowohl die Dauer als auch Anzahl, verhalten sich vergleichbar und weisen ähnliche Effekte auf.

Zusammenfassend finden sich in der Fahrprofilanalyse damit mehr langfristige Effekte als sie in den Analysen des Hauptdesigns gefunden wurden. Da die Teilnahme über alle vier Messzeitpunkte ein hohes Maß an Einsatzbereitschaft voraussetzte und nur eine vergleichsweise geringe Anzahl von Probanden an T4 noch teilnahmen, kann nicht ausgeschlossen werden, dass es sich bei diesen Befunden, um den Effekt einer Positivselektion handelt.

Für den letzten Bereich der Verhaltensebene – den *Selbst- und Fremdeinschätzungen der Einsatzfahrten* – konnten, wie auch im Hauptdesign, weder in kontrollierten noch in unkontrollierten Analysen Veränderungen zwischen den Zeitpunkten gefunden werden.

#### 4.4.3 Verhaltensänderung durch das Training in der ungepaarten Stichprobe

Wie bereits im Abschnitt 4.3.3 beschrieben, sollen auch für die Verhaltensebenen aufgrund der hohen Dropout-Quote noch einmal alle Daten in einem Design mit den Messzeitpunkten als unabhängigen Faktor

analysiert werden Erneut werden nur die Variablen in die Tabellen aufgenommen, die bereits Effekte gezeigt haben oder bei denen sich Effekte durch Hinzunahme der Dropouts ändern. Zusätzlich werden in den Berechnungen der Verhaltensvariablen nur die Gruppen 1 und 2 betrachtet, die auch tatsächlich zu beiden Zeitpunkten gemessen werden konnten. Da die Gruppe 3 nicht hinsichtlich der Anfangswerte bewertet werden kann, können Interaktionseffekte zwischen den Zeitpunkten nicht gefunden werden, diese Gruppe wird daher ausgeschlossen.

In Tabelle 39 werden zunächst die kurzfristigen Haupteffekte dargestellt, also die Unterschiede zwischen den Gruppen zu den Zeitpunkten T1 und T2. Hier konnten im Längsschnitt lediglich für die Geschwindigkeiten ein Trainingseffekt gefunden werden, dahingehend, dass die mittlere und maximale Geschwindigkeit in den Trainingsgruppen nach dem Training sinkt. Die maximale Längsbeschleunigung sank insgesamt bei allen Gruppen und in der eigenen Beurteilung der hektischen Fahrweisen gab es Unterschiede zwischen den Gruppen. Weitere Effekte konnten auf der Verhaltensebene hinsichtlich des Trainings nicht gefunden werden.

Die Effekte der mittleren und maximalen Geschwindigkeit werden durch die Hinzunahme der nicht gepaarten Werte nicht mehr signifikant, obwohl die deskriptiven Unterschiede in derselben Systematik wie in den Hauptanalysen bestehen bleiben (vgl. Abbildung 44). Denkbar ist das die mit Aufnahme der Dropout-Daten extrem erhöhten Ausgangswerte der Gruppe 1 hier die Veränderungen durch das Training verdecken.

Tabelle 39: Verhaltensebene: Kurzfristige Effekte in der ungepaarten Stichprobe

Abhängige Variable	N <sub>T1</sub> /N <sub>T2</sub>	Gruppen- effekt		Effekt des Zeitpunkts		Interaktions- effekt	
		F	p	F	p	F	p
Durchschnittliche Geschwindigkeit (in km/h)	140/121	3.05	.082 <sup>+</sup>	0.29	.593	2.01	.158
Maximale Geschwindigkeit (in km/h)	140/121	0.01	.909	1.11	.294	2.00	.159
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	140/121	0.00	.987	10.99	.001***	2.59	.109
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	140/121	0.59	.444	10.56	.001***	0.31	.581
Dauer Trigger positive Längsbeschleunigung (in 0.05sek/min) (5,9 m/s <sup>2</sup> )	140/121	1.10	.295	3.91	.049*	0.57	.452
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	140/121	1.51	.221	3.17	.076 <sup>+</sup>	0.93	.335
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	140/121	1.12	.291	3.64	.058 <sup>+</sup>	0.31	.577
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	140/121	0.41	.522	6.59	.011*	1.15	.284
Dauer Trigger negative Längsbeschleunigung (in 0.05sek/min) (3 m/s <sup>2</sup> )	140/121	0.67	.412	7.34	.007**	0.53	.468
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	140/121	1.00	.318	6.56	.011*	0.57	.452
Anzahl Trigger pro Minute negative Längsbeschleunigung (5,9 m/s <sup>2</sup> )	140/121	1.93	.166	3.07	.081 <sup>+</sup>	1.52	.219
Maximale positive Querbeschleunigung (in m/s <sup>2</sup> )	140/121	0.23	.632	12.10	.001***	0.05	.828
Mittlere positive Querbeschleunigung (in m/s <sup>2</sup> )	140/121	0.18	.676	8.97	.003**	0.02	.892
Maximale negative Querbeschleunigung (in m/s <sup>2</sup> )	140/121	0.02	.878	6.62	.011*	0.17	.679
Mittlere negative Querbeschleunigung (in m/s <sup>2</sup> )	140/121	0.03	.863	6.38	.012*	0.01	.926

Abhängige Variable	N <sub>T1</sub> /N <sub>T2</sub>	Gruppen- effekt		Effekt des Zeitpunkts		Interaktions- effekt	
		F	p	F	p	F	p
Dauer Trigger Querbeschleunigung (in 0.05sek/min) (6,9 m/s <sup>2</sup> )	140/121	1.72	.191	7.53	.006**	0.92	.339
Anzahl Trigger pro Minute Querbeschleunigung (6,9 m/s <sup>2</sup> )	140/121	1.62	.205	7.88	.005**	0.79	.375
Fahrerbeurteilung rasante Fahrweise	155/144	3.77	.053 <sup>+</sup>	0.04	.832	0.41	.521
Fahrerbeurteilung hektische Fahrweise	155/144	5.53	.019*	2.00	.158	0.01	.934

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht) mithilfe der gepaarten Stichprobe über die Zeitpunkte T1 und T2 zwischen den Gruppen 1 und 2 und Zeitpunkten; Abkürzungen: N=Anzahl der Probanden zu den zwei Zeitpunkten; F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06) Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

Bei allen Beschleunigungswerten inklusive der Trigger treten in der ungepaarten Stichprobe Effekte des Zeitpunkts in die Richtung auf, dass zu T2 geringere Beschleunigungswerte vorliegen als zu T1. Diese Effekte waren in den Hauptanalysen nicht zu sehen und könnten, ähnlich wie in den Geschwindigkeiten ein Hinweis darauf sein, dass zwischen T1 und T2 Probanden die Stichprobe verlassen haben, die „rasanter“ fahren, da diese die Werte zu T1 deutlich erhöhen und damit die Unterschiede zwischen den Zeitpunkten künstlich verstärken würden. Damit könnte auch dieser Aspekte ein Hinweis auf eine Positivselektion in der Stichprobe sein.

Neben den Fahrprofilaten bleibt der im Hauptdesign gefundene Gruppeneffekt der Selbstbeurteilung der hektischen Fahrweise auch unter Berücksichtig der Dropouts signifikant. Hinzu kommt ein tendenzieller Gruppenunterschied für die eigene Beurteilung der rasanten Fahrweise, in der sich die Gruppe 1 höher bewertet als Gruppe 2. Beide Effekte zeigen aber keinen Hinweis darauf, dass sie auf das Training zurückzuführen sind.

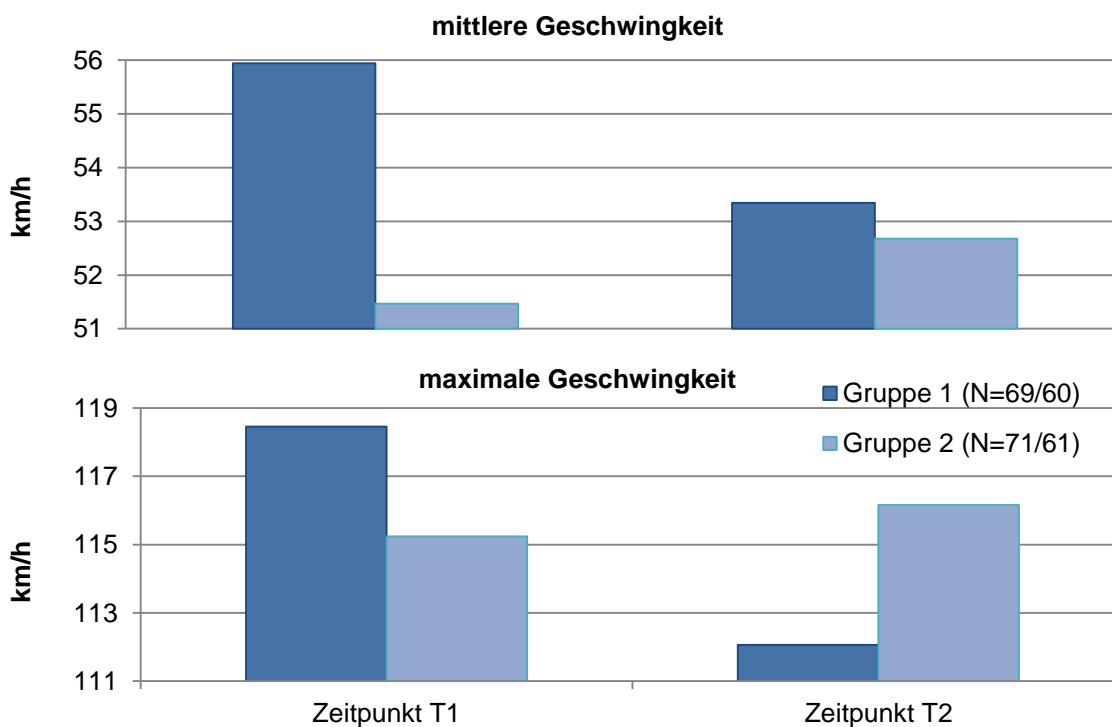


Abbildung 44: Mittlere und maximale Geschwindigkeit bei RTW-Fahrten zum Einsatzort im kurzfristigen Verlauf unter Hinzunahme der Dropouts



Im langfristigen Verlauf der Verhaltensebene konnten unter Kontrolle von Alter, Geschlecht und der Gruppe insgesamt sinkende Werte für die Geschwindigkeiten, die mittlere positive Längsbeschleunigung sowie die maximale Bremsverzögerung gefunden werden. Diese werden im Folgenden auch unter Berücksichtigung der Dropouts analysiert.

Tabelle 40: Verhaltensebene: Langfristige Effekte in der ungepaarten Stichprobe

Abhängige Variable	N <sub>T1</sub> /N <sub>T4</sub>	Gruppeneffekt		Effekt des Zeitpunkts		Interaktions-effekt	
		F	p	F	p	F	p
Durchschnittliche Geschwindigkeit (in km/h)	140/30	0.65	.420	1.23	.269	1.44	.232
Maximale Geschwindigkeit (in km/h)	140/30	2.01	.158	0.92	.338	0.50	.479
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	140/30	0.27	.604	1.19	.278	0.05	.825
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	140/30	0.00	.983	0.00	.945	1.15	.286
Mittlere negative Querbeschleunigung (in m/s <sup>2</sup> )	140/30	0.31	.578	4.35	.038*	0.44	.510
Fahrerbeurteilung rasante Fahrweise	155/36	5.37	.022*	0.15	.695	0.61	.437
Beifahrerbeurteilung hektische Fahrweise	146/30	4.57	.034*	0.08	.773	3.68	.057 <sup>+</sup>

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht) mithilfe der gepaarten Stichprobe über die Zeitpunkte T1 und T4 zwischen den Gruppen 1 und 2 und Zeitpunkten; Abkürzungen: N=Anzahl der Probanden zu den zwei Zeitpunkten; F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06) Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

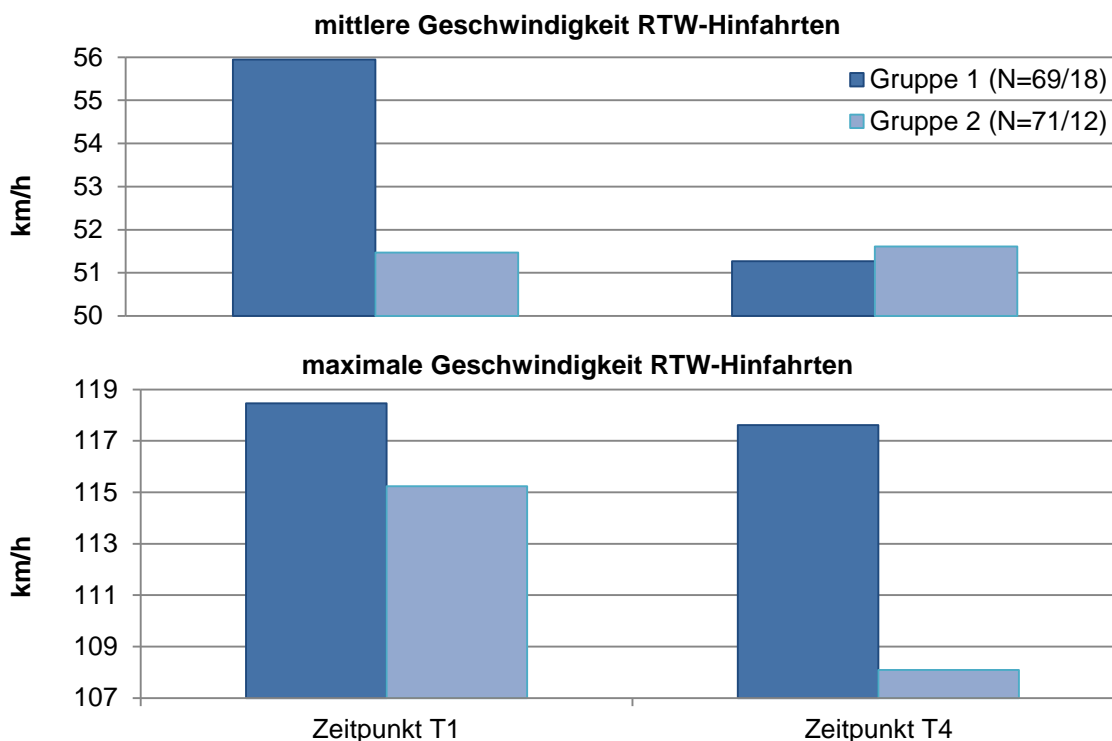


Abbildung 45: Mittlere und maximale Geschwindigkeit bei RTW-Fahrten zum Einsatzort im langfristigen Vergleich unter Hinzunahme der Dropouts

Wie bei den kurzfristigen Trainingseffekten, verschwindet durch die ungepaarte Analyse die *Geschwindigkeitsreduzierung* durch möglicherweise vergleichbare Gründe. Allerdings ist auch hier der in der Abbildung 45 gezeigte Trend eindeutig – in der Gruppe 2 verändert sich die mittlere Geschwindigkeit nicht, während die Teilnehmer der Gruppe 1 zu T4 über 4.5 km/h langsamer fahren als zu Beginn der Studie. Die maximale Geschwindigkeit wird in beiden Gruppen um durchschnittlich 3 km/h reduziert.

Die Effekte der *Längsbeschleunigung* verschwinden, wenn alle Probanden aufgenommen werden. Dafür zeigen sich erneut, wie auch schon bei den kurzfristigen Effekten - Effekte in der *Querb beschleunigung* zwischen den Zeitpunkten hinzu, die sich in beiden Gruppen ohne Unterschied zwischen den Gruppen verringern.

Für die *Einschätzung der Fahrweise* finden sich bei den ungepaarten Vergleichen, Gruppenunterschiede für die eigene Einschätzung der Rasanz, in der sich die Gruppe 1 insgesamt höher bewertet als die Gruppe 2. Die Beifahrer bewerten die hektische Fahrweise der Einsatzfahrer in der Gruppe 1 insgesamt höher und diese steigt langfristig an, wohingegen die der Gruppe 2 geringer bewertet wird. Dieser Interaktionseffekt wurde in der gepaarten Stichprobe nicht gefunden.

Für die weiteren Effekte des Trainings auf der Verhaltensebene fanden sich zusammenfassend folgende Befunde:

- Die Befunde zur mittleren und maximalen Geschwindigkeiten bei Einsatzfahrten sind im Trend, wenn auch nicht signifikant, in der Wartegruppe replizierbar.
- Die Befunde zu den Geschwindigkeiten sind nachhaltig und zeigen eine langfristig signifikante Verringerung der mittleren und maximalen Geschwindigkeiten.
- Zusätzlich finden sich langfristig signifikante geringere positive und negative Längsbeschleunigungen.
- Unter Hinzunahme der Dropout-Daten sind die Effekte auf die Fahrprofile nicht mehr zu sehen, was ein Hinweis darauf sein könnte, dass bestimmte Typen von Einsatzfahrern die Studie vorzeitig verlassen haben.

## 4.5 Weitere Ergebnisse zur Resultatebene

### 4.5.1 Veränderung der Resultate durch das Training in der Kontrollgruppe

Im folgenden Abschnitt wird die Replizierbarkeit der Ergebnisse für die Variablen der Resultatebene betrachtet. Die Daten für die Wartegruppe, die das Training zwischen T2 und T3 erhielten werden in Tabelle 41 dargestellt, wobei vor allem die Veränderungen zwischen diesen Zeitpunkten im Vergleich zu den Befunden in der Interventionsgruppe zwischen T1 und T2 von Interesse sind.

Tabelle 41: Resultateebene: Zeitlicher Verlauf über die ersten drei Messzeitpunkte in der Wartegruppe

Abhängige Variable	N	Zeitpunkt T1 (a)		Zeitpunkt T2 (b)		Zeitpunkt T3 (c)		Kontrollierter Zeiteffekt		Paarweise Vergleiche	
		M	SD	M	SD	M	SD	F	p	F	p
<b>Längerfristige Beanspruchungsfolgen – Irritation</b>											
Kognitive Irritation	64	2.48 <sup>(b),c</sup>	1.15	2.27 <sup>(a)</sup>	1.03	2.09 <sup>a</sup>	1.06	2.97	.055 <sup>+</sup>	<b>7.29</b>	.001 <sup>***</sup>
Emotionale Irritation	64	2.50 <sup>(b),c</sup>	1.02	2.27 <sup>(a)</sup>	0.95	2.21 <sup>a</sup>	1.08	0.93	.396	<b>4.28</b>	.018 <sup>*</sup>
Irritation gesamt	64	2.49 <sup>(b),c</sup>	0.96	2.27 <sup>(a)</sup>	0.91	2.17 <sup>a</sup>	0.99	2.10	.128	<b>7.22</b>	.002 <sup>**</sup>

Abschlussbericht „Achtung Blaulicht“

Abhängige Variable	N	Zeitpunkt T1 (a)		Zeitpunkt T2 (b)		Zeitpunkt T3 (c)		Kontrollierter Zeiteffekt		Paarweise Vergleiche	
		M	SD	M	SD	M	SD	F	p	F	p
<b>Längerfristige Beanspruchungsfolgen – körperliche Beschwerden</b>											
GBB Erschöpfung	63	4.83 <sup>c</sup>	3.79	4.49 <sup>(c)</sup>	3.83	3.81 <sup>a,(b)</sup>	4.16	0.25	.777	<b>4.98</b>	.010**
GBB Magenbeschwerden	63	2.86 <sup>c</sup>	4.25	2.40	3.99	2.05 <sup>a</sup>	3.56	0.07	.932	<b>4.41</b>	.016*
GBB Gliederschmerzen	64	5.25 <sup>(b),(c)</sup>	4.02	4.66 <sup>(a)</sup>	4.17	4.52 <sup>(a)</sup>	4.41	0.13	.849	<b>3.05</b>	.055 <sup>+</sup>
GBB Herzbeschwerden	61	1.70	3.65	1.56	3.63	1.21	3.35	0.51	.603	1.74	.184
GBB Beschwerdedruck	59	14.69 <sup>c</sup>	13.96	13.34	14.25	11.86 <sup>a</sup>	14.10	0.12	.887	<b>4.38</b>	.017*
<b>Längerfristige Beanspruchungsfolgen – Wohlbefinden als Depressionscreening und Burnout</b>											
Wohlbefinden (WHO-5)	64	3.23 <sup>(b)</sup>	0.92	3.44 <sup>(a)</sup>	0.80	3.33	0.87	0.13	.877	<b>2.75</b>	.072 <sup>+</sup>
Burnout (emotionale Erschöpfung)	64	2.57	1.09	2.56	1.04	2.57	1.15	0.01	.993	0.01	.991
Burnout (Zynismus)	64	2.52	1.02	2.53	1.12	2.66	1.16	1.28	.281	1.83	.169
<b>Schichtbezogene Beanspruchung – Veränderung der Befindlichkeit über die Schichtmessungen</b>											
Vitalität (BBS)	65	-0.80 <sup>b</sup>	0.81	-0.33 <sup>a,c</sup>	0.46	-0.66 <sup>b</sup>	0.71	1.32	.271	<b>10.64</b>	.000***
Gleichgewicht (BBS)	65	-0.27	0.48	-0.38	0.55	-0.30	0.46	0.36	.698	1.34	.270
Extraversion (BBS)	65	-0.24	0.59	-0.28	0.86	-0.20	0.51	1.67	.192	0.28	.756
Vigilanz (BBS)	65	-0.57	0.72	-0.70	0.90	-0.51	0.72	0.08	.910	1.57	.216
<b>Schichtbezogene Beanspruchung – Workload zum Ende der Schichtmessungen</b>											
NASA geistige Anforderungen	64	35.72 <sup>b,c</sup>	17.31	30.91 <sup>a</sup>	18.80	28.88 <sup>a</sup>	20.77	0.01	.982	<b>4.95</b>	.010**
NASA körperliche Anforderungen	64	33.98	16.00	31.88	19.91	29.02	20.14	0.10	.909	<b>2.09</b>	.132
NASA zeitliche Anforderungen	64	40.54	16.77	39.71	20.25	34.68	21.82	0.47	.627	<b>2.49</b>	.091 <sup>+</sup>
NASA Leistung	64	83.19	10.94	81.96	13.53	84.42	11.16	0.57	.569	1.61	.209
NASA Anstrengung	64	32.53 <sup>c</sup>	15.12	30.08	17.84	25.87 <sup>a</sup>	18.05	0.40	.670	<b>3.72</b>	.030*
NASA Frustration	64	21.51 <sup>c</sup>	16.33	21.23 <sup>c</sup>	15.88	16.31 <sup>a,b</sup>	13.56	0.28	.732	<b>3.68</b>	.031*
<b>Schichtbezogene Beanspruchung – Müdigkeit über die Schichtmessungen</b>											
Mittlere Müdigkeit (KSS)	65	3.40	1.16	3.34	1.28	3.30	1.22	0.53	.591	0.48	.621
Mittlere Ermüdung (KSS)	62	0.53	1.91	0.97	1.77	0.68	1.53	0.07	.932	1.29	.283
<b>Schichtbezogene Beanspruchung – kardiologische Werte während der Schichtmessungen</b>											
Mittlere Herzfrequenz (Schläge pro Minute)	39	91.95	9.17	90.74	9.33	91.10	9.21	1.07	.349	0.44	.651
Maximale Herzfrequenz (Schläge pro Minute)	39	142.05	16.95	138.77	13.54	139.23	11.37	0.62	.543	0.54	.588
Minimale Herzfrequenz (Schläge pro Minute)	39	56.13	8.35	56.67	7.72	57.10	7.37	0.17	.847	0.33	.720
<b>Typische Kennzahlen – Dauer der Einsatzfahrten</b>											
„Hilfsfrist“ (Dauer der Einsatzfahrten, in mm:ss)	57	6.50	2.23	6.95	2.52	6.34	2.43	0.69	.505	1.51	.230

Berechnung: Varianzanalyse mit Messwiederholung unter Kontrolle von Alter und Geschlecht; Abkürzungen: M=Mittelwert, SE=Standardfehler, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; Kontraste zur Berechnung der Zeiteffekte: einfach (T3 zu T1 und T3 zu T2) und umgekehrter Helmert (T3 zu Mittelwert aus T1 und T2); Bonferroni-Post-Hoc-Test: a-c kennzeichnet signifikante Unterschiede (p≤0.05), (a)-(c) kennzeichnet tendenziell signifikante Unterschiede (p≤0.10) zwischen den Zeitpunkten

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

Insgesamt gab es für die Interventionsgruppe im Bereich der Resulateebene in der Hauptanalyse keine eindeutig auf das Training zurückführbare Effekte. In der Wartegruppe sieht dies leicht anders aus.

Bei den *längerfristigen Beanspruchungsfolgen* findet sich ein Haupteffekt der Zeit für die kognitive Irritation, die in der Wartegruppe zwischen T1 und T2 leicht sinkt und dann eher konstant bleibt, also keinen zusätzlichen Trainingseffekt zeigt. Allerdings findet sich im kontrollierten Design ein Interaktionseffekt mit dem Geschlecht über die Zeitpunkte hinweg ( $F_{(2,122)}=3.42$ ,  $p=.036$ ), der, wie in Abbildung 46, dargestellt, erst zwischen den Zeitpunkten T2 und T3 auftritt, also zum Zeitpunkt des Trainings. Es hat also den Anschein, dass es hier eine differenzielle Wirkung des Trainings geben könnte, allerdings wurde dieser Zusammenhang bei den Hauptanalysen gefunden, so dass die Annahme von Kausalität schwierig ist.

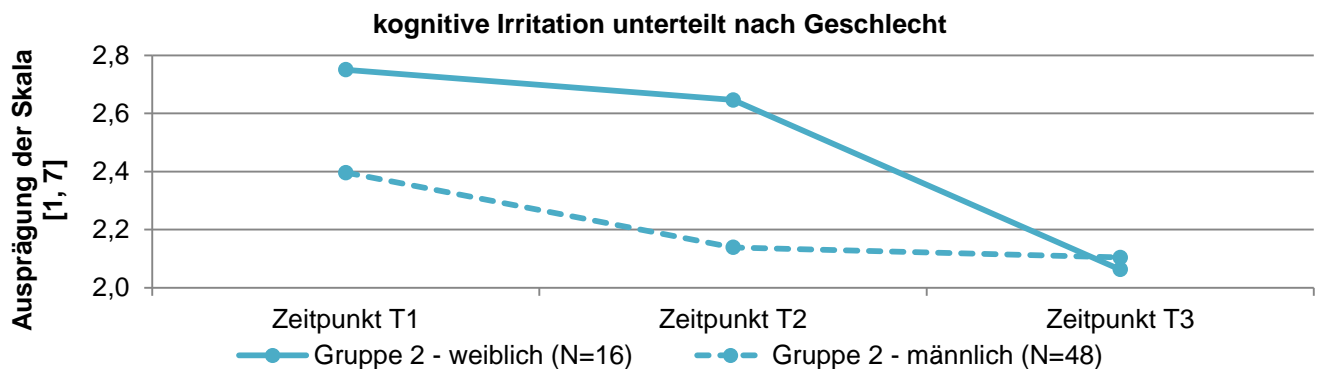


Abbildung 46: Kognitive Irritation in Abhängigkeit vom Geschlecht

Bei den körperlichen Beschwerden als Parameter der längerfristigen Beanspruchungsfolgen zeigte die Stichprobe insgesamt erheblich erhöhte Werte, ohne dass das Training darauf einen Einfluss gehabt hätte. Auch in der Gruppe 2 findet sich kein Gesamtzeiteffekt über die drei Zeitpunkte. Allerdings können in den paarweisen Vergleichen einige Unterschiede zwischen den Zeitpunkten gefunden werden. Unterschiede im Trainingszeitraum (T2-T3) finden sich allerdings ausschließlich für die Erschöpfungsneigung. Der Anteil von Personen mit kritischen Werten sinkt in diesem Zeitraum von 43.8% auf 35%. Dies ist allerdings eingebettet in einen Trend der auch für Magenbeschwerden, Gliederschmerzen und den Beschwerdedruck bereits zwischen T1 und T2 in der Wartegruppe sinkende Werte zeigt, so dass ein Trainingseffekt hier nur sehr bedingt angenommen werden kann ist.

Für die Beanspruchungsfolgen Burnout und Wohlbefinden/Depression konnten ebenso wie in den Interventionsgruppen auch in der Wartegruppe kein Hinweis auf Trainingseffekte gefunden werden. Die Ausprägung des Zynismus als Burnoutkomponente ist in Gruppe 2 weiterhin abhängig vom Geschlecht ( $F_{(1,61)}=4.67$ ,  $p=.035$ ) nicht jedoch vom Alter. Die Männer bewerten sich selbst zu allen drei Zeitpunkten als zynischer als die Frauen der Wartegruppe.

Bei den *schichtbezogenen Beanspruchungsmaßen* sind in der Wartegruppe keinerlei Effekte der Zeit festzustellen, allerdings treten für Befindlichkeit und Workload einige signifikante Unterschiede zwischen einzelnen Messzeitpunkten auf. Die Veränderung der Vitalität in der Gruppe 2 über die Schichten zeigt einen signifikanten Unterschied des Zeitpunktes T2 im Vergleich zu den anderen beiden Zeitpunkten. Da diese Veränderung auch schon zwischen T1 und T2 gefunden wird und zu T3 wieder auf das Ausgangsniveau von T1 zurückgeht, kann trotz dieses Unterschiedes nicht von einem Trainingseffekt ausgegangen werden. Ähnlich wie bei den anderen Analysen zeigen die Kontrollvariablen zum Teil Effekte auf die Befindlichkeitsänderungen während der Schichten. So findet sich ein tendenzieller Alterseinfluss auf die Veränderung der Vitalität ( $F_{(1,62)}=2.81$ ,  $p=.099$ ) und Vigilanz ( $F_{(1,62)}=2.87$ ,  $p=.095$ ) über die Schichten sowie ein tendenzieller Interaktionseffekt des Geschlechts mit der Zeit für die soziale Extravertiertheit

( $F_{(2,124)}=2.70$ ,  $p=.071$ ). Bei den jüngeren Probanden gibt es eine größere Differenz der Vitalität und Vigilanz zwischen Schichtende und Schichtbeginn als für die älteren Probanden.

Für den Workload als schichtbezogenem Beanspruchungsmaß finden sich – wie schon bei der Befindlichkeit – keinerlei generellen Zeiteffekt in der Wartegruppe. Allerdings treten deutliche Unterschiede zwischen T3 und den beiden ersten Messzeitpunkten im Bereich der Frustration auf, die in der Wartegruppe zwischen T2 und T3 deutlich sinkt. Für die geistigen Anforderungen finden sich Unterschiede zwischen T1 und den beiden folgenden Zeitpunkten und für die Anstrengung ein Unterschied zwischen T1 und T3 (vgl. Abbildung 47). Alter und Geschlecht haben als Kontrollvariablen keinen Einfluss auf den berichteten Workload der Wartegruppe. Insgesamt erscheint damit nur bei der Frustration am Ende der Schicht ein potenzieller Trainingseffekt zu bestehen, der allerdings nicht in den Hauptanalysen zu sehen war, und damit vermutlich wenig belastbar ist.

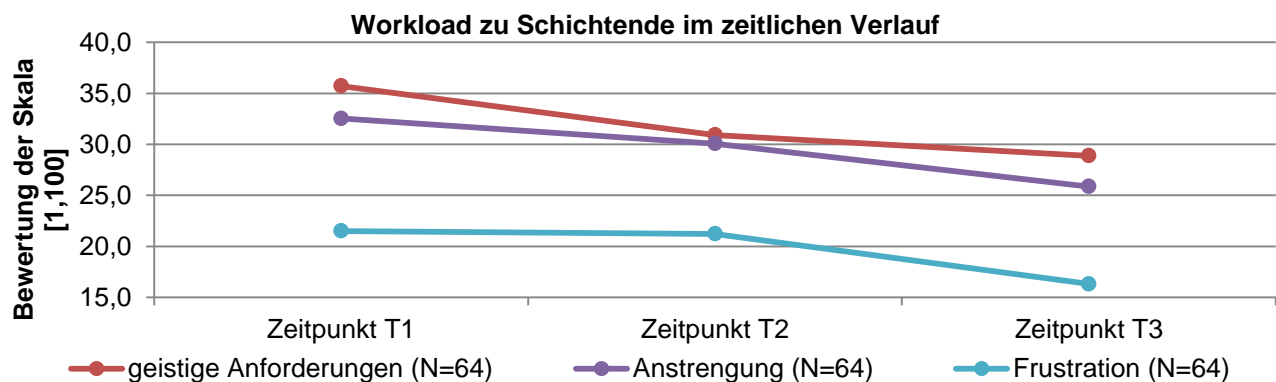


Abbildung 47: Geistige Anforderungen, Anstrengung und Frustration zum Schichtende im zeitlichen Verlauf für die Wartegruppe

Hinsichtlich der schichtbezogenen Müdigkeit und Ermüdung kann wie auch schon in den Hauptanalysen kein Trainingseffekt gefunden werden. Es gibt weder einen kontrollierten Zeiteffekt noch Unterschiede zwischen den einzelnen Messzeitpunkten. Bezogen auf die Kontrollvariablen fand sich im Hauptdesign ein Effekt des Geschlechts, der sich in der Wartegruppe allein nicht mehr zeigt. Dafür gibt es in der Wartegruppe einen Einfluss des Alters auf die Müdigkeit zum Schichtende ( $F_{(1,62)}=6.58$ ,  $p=.013$ ). Die jüngeren Probanden geben am Ende der erfassten Schichten eine höhere Müdigkeit an als ihre älteren Kollegen.

Dasselbe Ergebnis zeigt sich auch für die minimale, maximale und mittlere Herzfrequenz während der Schichten. Wie schon im Vergleich der Interventions- mit der Wartegruppe, zeigen sich in der Wartegruppe keinerlei Veränderungen über die Zeit insgesamt oder durch das Training zwischen T2 und T3. Hinsichtlich der Kontrollvariablen findet sich für das Alter ein Einfluss auf die maximale Herzfrequenz in der Wartegruppe ( $F_{(1,36)}=5.07$ ,  $p=.031$ ). Hier haben die jüngeren Probanden insgesamt höhere maximale Herzfrequenzen als ihre älteren Kollegen. In den Hauptauswertungen wurde für die jüngeren Teilnehmer eine geringere minimale Herzfrequenz gefunden, was insgesamt einen breiteren Streubereich der Werte zeigt und möglicherweise auf eine bessere Anpassungsfähigkeit des kardiovaskulären Systems bei jüngeren Teilnehmern hindeutet. Insgesamt zeigt sich aber auch für diesen Bereich kein Effekt des Trainings in der Warte- wie in der Interventionsgruppe.

Auch im letzten erfassten Bereich der Resultateebene – der „Hilfsfrist“ – können keinerlei relevante Effekte durch das Training gefunden werden. Wie schon dargelegt, sollte das Training für diesen Bereich nicht zu einer unnötigen Verlängerung der Einsatzfahrten führen.

## 4.5.2 Veränderung der Resultate im langfristigen Verlauf

Auch für die Resultateebene wird der längerfristige Erfolg des Trainings anhand eines Vergleichs zwischen Follow-up-Messungen (T4) und den ersten Messungen (T1) unter Kontrolle von Gruppe, Alter und Geschlecht überprüft (vgl. Tabelle 42). Anders als bei den Variablen der vorherigen Evaluationsebenen, wird gerade bei den Resultaten erwartet, dass die Auswirkungen des Trainings hier eher im längerfristigen Verlauf zu sehen sind als unmittelbar nach dem Training.

Insgesamt finden sich auch im längerfristigen Vergleich allerdings nur wenige signifikante bzw. Effekte mit mittlerer Effektstärke, wenn neben dem Alter und Geschlecht auch für die Studiengruppe kontrolliert wird. In unkontrollierten Analysen sind erheblich mehr Unterschiede zu finden, was erneut dafür spricht, dass sich die drei untersuchten Gruppen nicht parallel entwickeln.

Tabelle 42: Resultateebene: Langfristiger Verlauf in der Gesamtstichprobe

Abhängige Variable über alle Gruppen	N	Zeitpunkt T1 (a)		Zeitpunkt T4 (b)		Zeiteffekt	
		M	SD	M	SD	F	P
<b>Längerfristige Beanspruchungsfolgen – Irritation</b>							
Kognitive Irritation (*)	45	3.05	1.29	2.61	1.53	0.07	.790
Emotionale Irritation (*)	45	2.85	1.09	2.50	1.21	0.34	.565
Irritation gesamt (*)	45	2.93	1.07	2.54	1.28	0.07	.791
<b>Längerfristige Beanspruchungsfolgen – körperliche Beschwerden</b>							
GBB Erschöpfung	43	5.42	4.90	5.70	5.17	0.92	.345
GBB Magenbeschwerden	45	3.49	4.27	3.04	4.18	0.14	.712
GBB Gliederschmerzen	44	6.91	5.10	5.89	5.17	0.80	.375
GBB Herzbeschwerden (*)	45	2.76	4.13	2.40	4.33	0.53	.472
GBB Beschwerdedruck	43	18.47	16.48	17.16	16.42	0.81	.373
<b>Längerfristige Beanspruchungsfolgen – Wohlbefinden als Depressionsscreening und Burnout</b>							
Wohlbefinden (WHO-5)	44	3.18	0.95	2.96	1.16	1.39	.246
Burnout (emotionale Erschöpfung)	44	2.69	1.21	2.96	1.52	0.34	.561
Burnout (Zynismus)	44	2.74	1.22	2.57	1.15	1.26	.268
<b>Schichtbezogene Beanspruchung – Veränderung der Befindlichkeit über die Schichtmessungen</b>							
Vitalität (BBS)	34	-0.69	0.80	-0.51	0.57	<b>2.35</b>	.136
Gleichgewicht (BBS) (*)	34	-0.20	0.45	-0.49	0.57	0.15	.699
Extraversion (BBS)	34	-0.29	0.64	-0.36	0.80	0.48	.492
Vigilanz (BBS)	34	-0.58	0.60	-0.60	0.73	<b>2.13</b>	.155
<b>Schichtbezogene Beanspruchung – Workload zum Ende der Schichtmessungen</b>							
NASA geistige Anforderungen (*)	37	43.17	17.24	37.13	27.59	<b>3.38</b>	.075 <sup>+</sup>
NASA körperliche Anforderungen	37	37.68	16.24	36.43	25.32	1.78	.191
NASA zeitliche Anforderungen	37	46.55	18.35	41.97	25.79	<b>2.48</b>	.125
NASA Leistung	37	81.98	12.49	82.11	13.53	0.34	.563
NASA Anstrengung	37	35.35	16.48	34.22	25.24	0.47	.500
NASA Frustration	37	25.30	21.63	23.20	21.70	<b>2.79</b>	.104
<b>Schichtbezogene Beanspruchung – Müdigkeit über die Schichtmessungen</b>							
Mittlere Müdigkeit (KSS)	37	3.41	1.37	3.48	1.33	1.52	.227
Mittlere Ermüdung (KSS) (*)	37	0.50	1.98	1.10	1.84	0.07	.787
<b>Schichtbezogene Beanspruchung – kardiologische Werte während der Schichtmessungen</b>							
Mittlere Herzfrequenz (Schläge pro Minute) (*)	18	88.06	11.14	84.61	11.37	<b>2.38</b>	.145
Maximale Herzfrequenz (Schläge pro	18	130.78	12.73	132.11	16.08	0.03	.855

Abhängige Variable über alle Gruppen	N	Zeitpunkt T1 (a)		Zeitpunkt T4 (b)		Zeiteffekt	
		M	SD	M	SD	F	P
Minute)							
Minimale Herzfrequenz (Schläge pro Minute) (*)	18	57.50	9.37	53.67	9.55	<b>1.26</b>	.280
Typische Kennzahlen – Dauer der Einsatzfahrten							
„Hilfsfrist“ (Dauer der Einsatzfahrten, in mm:ss)	32	5.95	1.74	6.29	2.49	1.51	.229

Berechnung: Varianzanalyse (Kontrolle von Alter, Geschlecht und Versuchsgruppe); Abkürzungen: M=Mittelwert, SD=Standardabweichung, F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; (\*) hinter den Skalen kennzeichnet Unterschiede zwischen T1 und T4, die bei unkontrollierter Analyse mit mindestens 10% Irrtumswahrscheinlichkeit signifikant sind  
 Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Im Einzelnen zeigen sich für die *längerfristigen Beanspruchungsfolgen* im Bereich der Irritation, keine Effekte im kontrollierten T1-T4-Vergleich, allerdings treten hier signifikante Interaktionseffekte zwischen den Zeitpunkten und der Versuchsgruppe (kognitive Irritation:  $F_{(1,41)}=5.17$ ,  $p=.028$ ,  $\eta^2_{part}=.112$ , emotionale Irritation:  $F_{(1,41)}=3.89$ ,  $p=.055$ ,  $\eta^2_{part}=.087$ ; Gesamtirritation:  $F_{(1,41)}=5.64$ ,  $p=.022$ ,  $\eta^2_{part}=.121$ ) auf. Exemplarisch ist dies für die kognitive Irritation in Abbildung 48 dargestellt.

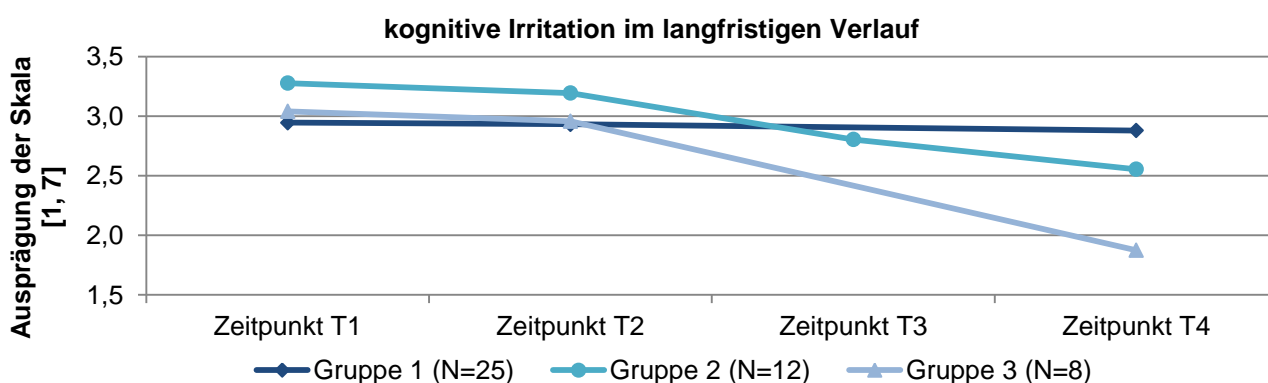


Abbildung 48: Kognitive und emotionale Irritation im langfristigen Verlauf unterteilt nach Versuchsgruppen

Auch für die körperlichen Beschwerden als Parameter der längerfristigen Beanspruchungsfolgen finden sich im langfristigen Vergleich unter Kontrolle von Alter, Geschlecht und Studiengruppe keine Unterschiede. Unkontrolliert sinken die Herzbeschwerden an T4 leicht ab, was sich in den alters- und geschlechtsadjustierten Prozenträngen in einer Veränderung von 42.2% Personen mit kritischen Werten zu T1 auf 33.3% zu .T4 darstellt. Hinsichtlich der Kontrollen finden sich erneut Effekte, die nicht vom Training abhängen: So ist die Erschöpfungsneigung tendenziell abhängig vom Geschlecht ( $F_{(1,39)}=3.02$ ,  $p=.090$ ,  $\eta^2_{part}=.072$ ) - Frauen geben durchgehend höhere Werte an als Männer. Magenbeschwerden ( $F_{(1,41)}=3.95$ ,  $p=.054$ ,  $\eta^2_{part}=.088$ ), Gliederschmerzen ( $F_{(1,40)}=5.18$ ,  $p=.028$ ,  $\eta^2_{part}=.115$ ) und der Gesamtbeschwerdedruck ( $F_{(1,39)}=4.11$ ,  $p=.049$ ,  $\eta^2_{part}=.095$ ) zeigen jeweils einen Alterseinfluss, in die erwartbare Richtung, dass ältere Probanden jeweils höhere Werte aufweisen als jüngere.

Wie bereits im kurzfristigen Verlauf, ändern sich auch im langfristigen Verlauf die Werte für das Wohlbefinden und Burnout nicht, weder unter Kontrolle von Alter, Geschlecht und Gruppe noch im direkten Mittelwertvergleich, so dass nicht von einem Effekt des Trainings ausgegangen werden kann. Alter und Geschlecht haben insgesamt auch keinen Einfluss auf das Wohlbefinden oder die Burnout-Komponente

emotionale Erschöpfung. Auf den Zynismus wirkt sich das Geschlecht in Interaktion mit der Zeit signifikant aus ( $F_{(1,40)}=4.50$ ,  $p=.040$ ,  $\eta^2_{\text{part}}=.101$ ), das Alter zeigt einen Gesamteffekt ( $F_{(1,40)}=3.56$ ,  $p=.066$ ,  $\eta^2_{\text{part}}=.082$ ). Die jüngeren Probanden zeigen insgesamt höheren Zynismus als die älteren Probanden. Der Zynismus bei den Männern nimmt leicht ab wohingegen er bei den Frauen leicht ansteigt über die Zeit.

Im Bereich der schichtbezogenen Beanspruchungen zeigen sich hinsichtlich der Befindlichkeit unter Kontrolle von Alter, Geschlecht und Gruppe keine signifikanten Unterschiede zwischen den ersten Messungen und den Follow-up-Messungen. Unkontrolliert ergibt sich ein Effekt in der Bewertung des intrapsychischen Gleichgewichtszustandes. Die Veränderung in diesen Werten wird zu T4 höher als sie zu T1 ist. Bei Vitalität und Vigilanz finden sich im kontrollierten Modell relativ hohe Effektstärken, ohne dass der Zeiteffekt signifikant wird. Dies ist vermutlich auf eine signifikante Interaktionseffekte zwischen den Zeitpunkten und dem Geschlecht zurückzuführen. Die Frauen weisen zu Beginn der Messungen höhere Differenzen über die Rettungsdienstschichten hinweg auf, als die Männer. Zu T4 kehrt sich dieser Unterschied jeweils um. Damit ist für die Befindlichkeit kein belastbarer langfristiger Effekt der Trainings festzustellen.

Hinsichtlich der Maße des Workloads zum Schichtende gibt es in den kontrollierten Analysen im langfristigen Verlauf eine signifikante Verringerung der wahrgenommenen geistigen Anforderungen. Dieser Effekt ist auch in den unkontrollierten Analysen vorhanden und wird in Abbildung 49 mit der Kontrollvariablen Alter illustriert. Alle anderen Maße zeigen keine systematischen Unterschiede zwischen den Zeitpunkten T1 und T4, obwohl sich relativ hohe Effektstärken für die wahrgenommenen zeitlichen Anforderungen und die Frustration finden. Diese und die weiteren NASA-TLX Aspekte weisen alle mindestens tendenzielle Interaktionseffekte zwischen den Zeitpunkten und dem Alter der Probanden auf, die nicht ohne weiteres erklärt werden können und möglicherweise mit Schichtspezifika zusammenhängen. Insgesamt können aber im Bereich des Workloads keine belastbaren langfristigen Effekte des Trainings identifiziert werden.

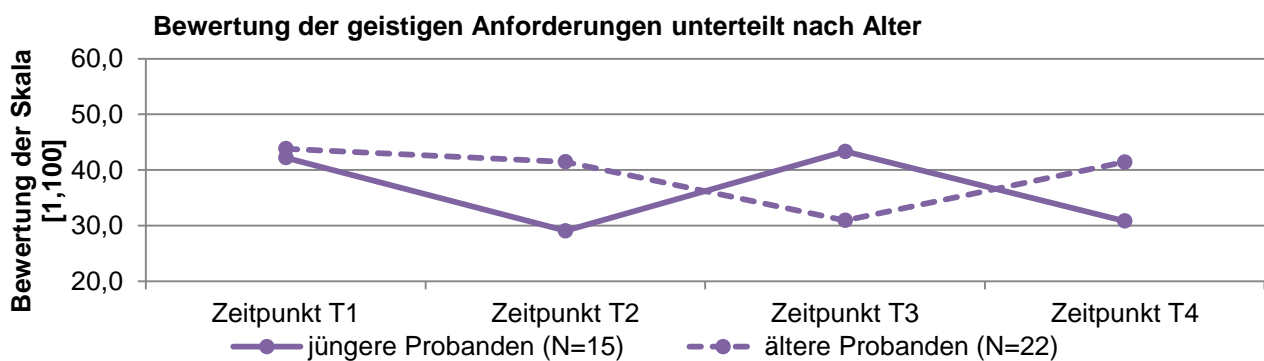


Abbildung 49: Geistige Anforderungen zum Schichtende in Abhängigkeit vom Alter im langzeitlichen Verlauf

Für Müdigkeit und Ermüdung als schichtbezogene Beanspruchungsmaße konnten in den Hauptanalysen keine Trainingseffekte gefunden werden. Auch im langfristigen Verlauf zwischen den ersten Messungen und den Follow-up-Messungen finden sich in den kontrollierten Analysen keine Veränderungen. Unkontrolliert zeigt sich jedoch ein tendenzieller Unterschied für die Ermüdung dahingehend, dass die Teilnehmer zum Zeitpunkt T4 eine höhere Ermüdung über die Schicht aufweisen als zu T1. Dieser Unterschied scheint bestimmt zu sein durch einen Interaktionseffekt zwischen den Zeitpunkten und dem Geschlecht der Probanden ( $F_{(1,33)}=4.62$ ,  $p=.039$ ,  $\eta^2_{\text{part}}=.123$ ). Die Frauen haben zu T1 eine etwas höhere Differenz, ermüden also etwas mehr als die Männer über die Schichten. Zu T4 weisen sie hingegen negative Werte auf, sind zum Schichtende also etwas wacher als zu Schichtbeginn. Bei den Männern steigt hingegen die Ermüdung über die Schichten zwischen T1 und T4 an. Auch in diesem Bereich lässt sich also kein langfristiger Trainingseffekt feststellen.



Ein leicht anderes Bild ergibt sich für die kardiologischen Werte während der Schichtmessung, bei denen sich in der Hauptanalyse keine Effekte zeigten: Unter Kontrolle von Alter, Geschlecht und Gruppe zeigen sich zwar auch keine Effekte, allerdings sind für die mittlere und minimale Herzfrequenz relativ hohe Effektstärken zu sehen, die sich auch in signifikanten Unterschieden bei den unkontrollierten Analysen widerspiegeln. Bei beiden Parametern sinken die Werte von T1 zu T4 in Abhängigkeit vom Alter ( $F_{(1,14)}=5.36$ ,  $p=.036$ ,  $\eta^2_{\text{part}}=.277$ ) – bei den jüngeren Teilnehmern nehmen die Werte stärker ab als bei den älteren. Insgesamt ergibt sich aber auch hier kein Hinweis darauf, dass sich das Training positiv auf die kardiologischen Werte ausgewirkt hat.

Für den letzten Parameter der Resultateebene, der „Hilfsfrist“ im Sinne der Dauer der Einsatzfahrten, zeigen sich auch im langfristigen Vergleich keinerlei Veränderungen.

### 4.5.3 Veränderung der Resultate durch das Training in der ungepaarten Stichprobe

Auch für die Resultateebene werden im Folgenden Analysen mit ungepaarten Stichproben, also unter Einschluss von Dropouts durchgeführt. Für die Daten, in denen zum Zeitpunkt T1 keine Daten in der Gruppe 3 vorliegen, werden auch nur die Gruppen 1 und 2 miteinander verglichen. Die langfristigen Daten, die in allen Gruppen auch vor dem Training erfasst wurden, werden über alle Gruppen hinweg analysiert.

In den gepaarten Hauptanalysen wurden keine Trainingseffekte auf der Resultateebene im kurzfristigen T1-T2 Vergleich sondern lediglich Gruppenunterschiede für die geistigen und körperlichen Anforderungen sowie ein leichter Anstieg der Dauer der Einsatzfahrten gefunden. Auch die in Tabelle 43 vorgestellten Analysen unter Einbezug der Dropouts zeigen keine Interaktionseffekte als Hinweise auf Einflüsse des Trainings, allerdings finden sich weitere Haupteffekte der Gruppen und/oder der Zeitpunkte.

Tabelle 43: Resultateebene: Kurzfristige Effekte in der ungepaarten Stichprobe

Abhängige Variable	N <sub>T1</sub> /N <sub>T2</sub>	Gruppeneffekt		Effekt des Zeitpunkts		Interaktionseffekt	
		F	p	F	p	F	p
Kognitive Irritation	179/165	2.91	.056 <sup>+</sup>	4.10	.044*	0.14	.870
GBB Gliederschmerzen	179/164	3.07	.048*	0.43	.514	0.04	.959
Wohlbefinden (WHO-5)	179/164	2.81	.061 <sup>+</sup>	0.15	.699	0.62	.536
Vitalität (BBS)	155/145	0.20	.654	32.94	.000***	0.36	.549
NASA geistige Anforderungen	153/142	5.05	.025*	3.21	.074 <sup>+</sup>	0.40	.529
NASA körperliche Anforderungen	153/142	6.54	.011*	0.23	.632	0.60	.439
NASA Anstrengung	152/142	4.21	.041*	2.62	.106	0.16	.692
Maximale Herzfrequenz (Schläge pro Minute)	135/116	0.59	.442	5.12	.024*	0.09	.763

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht) mithilfe der gepaarten Stichprobe über die Zeitpunkte T1 und T2 zwischen den Gruppen und Zeitpunkten; Abkürzungen: N=Anzahl der Probanden zu den zwei Zeitpunkten; F=F-Wert der Varianzanalyse, p=Signifikanzniveau: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles  $\eta^2 \geq .06$ )

Im Bereich der *längerfristigen Beanspruchungsfolgen* finden sich Effekte bei der kognitiven Irritation, den Gliederschmerzen und dem Wohlbefinden. Die kognitive Irritation ist insgesamt in der Gruppe 2 geringer

ausgeprägt, dieser Effekt war in den bisherigen gepaarten Analysen nur in den Prozenträgen sichtbar. Alle drei Gruppen sinken zwischen den Zeitpunkten T1 und T2, unabhängig vom Training. Bei den Glieder-schmerzen findet sich ein Gruppenunterschied, die Gruppe 3 gibt hier die höchsten Werte an, die Gruppe 2 die geringsten. Diese Werte ändern sich jedoch nicht über die Zeit. Hinsichtlich des Wohlbefindens gibt es ebenfalls einen Gruppenunterschied, die Gruppe 2 hat insgesamt die höchsten Werte. Dieser Trend war auch in den gepaarten Stichproben schon erkennbar.

Hinsichtlich der schichtbezogenen Beanspruchungen bleiben die Gruppenunterschiede bei den wahrge-nommenen geistigen und körperlichen Anforderungen auch unter Einbezug der Dropouts bestehen, wobei auch eine tendenzielle Abnahme bei den geistigen Anforderungen zu verzeichnen ist. Für die Anstrengungs-wahrnehmung zeigt sich in der ungepaarten Stichprobe ebenfalls ein Gruppeneffekt, in die Richtung, dass Gruppe 1 diese höher bewertet als Gruppe 2, und neu hinzu kommt eine hochsignifikante Verringerung der Vitalitätsveränderung über die Schichten in allen Gruppen. Schließlich findet sich ebenfalls als Effekt, der nur in der Stichprobe inklusive Dropouts auftritt, ein Absinken der maximalen Herzfrequenz beiden be-trachteten Gruppen zwischen T1 und T2.

Anders als der Zeiteffekt in den Hauptanalysen verändert sich die Dauer der Einsatzfahrten über die Zeit in der ungepaarten Stichprobe nicht mehr signifikant. Nur die absolute Höhe der Veränderung sinkt etwas.

Zusammengefasst zeigt die Hinzunahme der Dropouts auf der Resultateebene keine nennenswerten Unter-schiede zwischen den längsschnittlichen Hauptanalysen. Einen belastbaren Trainingseffekt gibt es kurz-fristig auf keiner der Reaktionsvariablen.

Für die langfristigen Veränderungen konnte in den Analysen der gepaarten Stichproben lediglich ein Zeit-effekt auf wahrgenommenen geistigen Anforderungen gefunden werden. Bei Irritation und mittlerer Herzfrequenz traten unerwartete Interaktionseffekte der zeitlichen Veränderungen mit der Versuchs-gruppe auf. Tabelle 44 zeigt die Änderungen dieser Effekte respektive neue Effekte, die bei Einbezug der Dropouts in ungepaarten Analysen zwischen T1 und T4 auftreten.

Tabelle 44: Resultateebene: Langfristige Effekte in der ungepaarten Stichprobe

Abhängige Variable	N <sub>T1</sub> /N <sub>T4</sub>	Gruppeneffekt		Effekt des Zeitpunkts		Interaktions-effekt	
		F	p	F	p	F	p
Kognitive Irritation	179/45	1.43	.242	2.87	.091 <sup>+</sup>	1.77	.173
Burnout (Zynismus)	179/44	1.40	.248	0.27	.602	2.41	.092 <sup>+</sup>
Vitalität (BBS)	155/34	0.36	.551	5.54	.020*	0.47	.492
NASA Anstrengung	152/37	3.64	.058 <sup>+</sup>	0.45	.504	1.30	.256
Mittlere Herzfrequenz (Schläge pro Minute)	135/17	1.51	.221	6.61	.011*	1.53	.218
Maximale Herzfrequenz (Schläge pro Minute)	135/17	1.36	.246	3.75	.055 <sup>+</sup>	0.32	.570
Minimale Herzfrequenz (Schläge pro Minute)	133/17	2.10	.149	3.52	.063 <sup>+</sup>	1.56	.214

Berechnung: Varianzanalyse (Kontrolle von Alter und Geschlecht) mithilfe der gepaarten Stichprobe über die Zeitpunkte T1 und T4 zwischen den Gruppen und Zeitpunkten; Abkürzungen: N=Anzahl der Probanden zu den zwei Zeitpunkten; F=F-Wert der Varianzanalyse, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Fett hervorgehobene F-Werte haben weisen mindestens eine mittlere Effektstärke auf (partielles η<sup>2</sup> ≥ .06)

Der auffälligste Unterschied ist, dass die langfristige Abnahme der wahrgenommenen geistigen Anfor-derungen verschwindet. Dies könnte – vorsichtig interpretiert – darauf hindeuten, dass Personen, die die

Studie verlassen haben, eher geringere Anforderungen in der T1-Messung angegeben haben, so dass sich die Wertdifferenzen insgesamt verringern. Die Ergebnisse der kognitiven Irritation verändern sich zwischen den verschiedenen Analysen nicht wesentlich: weiterhin sinkt sie über die Zeitpunkte unabhängig von den Gruppen. Neu zeigt sich unter Einbezug der Dropouts ein kleiner Interaktionseffekt zwischen Gruppe und Zeitpunkt hinsichtlich der Burnoutkomponente Zynismus. Gruppe 2 steigt über die drei Zeitpunkte in ihren Werten an, während Gruppe 1 sinkende Werte hat und Gruppe 3 keine Veränderungen aufweist.

Bei den schichtbezogenen Beanspruchungen zeigt sich im langfristigen Vergleich in den ungepaarten Stichproben zudem ein Absinken der Vitalitätsveränderung über die Schichten, wie dies auch in den Analysen zu kurzfristigen Effekten (T1 - T2) mit der ungepaarten Stichprobe gefunden wurde. Dies könnte – erneut mit aller Vorsicht - darauf hindeuten, dass vor allem Personen frühzeitig aus der Studie ausgeschieden sind, die während der Schichten stärker beansprucht waren (im Sinne höherer Vitalitätsverringernungen).

Für die wahrgenommene Anstrengung findet sich hingegen nur ein Gruppeneffekt, der ebenfalls auch in den T1-T2-Analysen der ungepaarten Stichprobe auftrat: Gruppe 1 wertet diese auch im langfristigen Vergleich höher als Gruppe 2. Auch dieser Effekt ist nicht bei den längsschnittlichen Analysen zu finden - allerdings bleibt fraglich, ob vergleichbar den Vitalitätsveränderungen Personen die Studie verlassen haben, die die Schichten als sehr anstrengend erlebt haben. Dies würde ein differenzielles „Ausscheiden“ aus Gruppe 1 voraussetzen, dass sich nicht völlig klären lässt (zwischen T1 und T2 haben 12 Personen die Gruppe 1 und 6 Personen die Gruppe 2 verlassen).

Für die Maße der Herzfrequenz finden sich bei den ungepaarten Analysen ausschließlich Effekte des Zeitpunkts – alle Werte sinken zwischen T1 und T4.

Zusammengefasst zeigt die Hinzunahme der Dropouts kaum wesentliche Veränderungen zu den gepaarten Analysen für langfristige Effekte. Allerdings scheinen sich gerade in den Parametern der Gesundheit/ Beanspruchung potenzielle Verzerrungen in der Stichprobe abzuzeichnen.

Für die weiteren Effekte des Trainings auf der Resultateebene fanden sich zusammenfassend folgende Befunde:

- Entgegen dem Fehlen von Effekten auf dieser Ebene im Hauptdesign fanden sich hier für die Wartegruppe nach dem Training einige Veränderungen, von denen jedoch viele bereits auf Veränderungen zwischen T1 und T2 basieren.
- Signifikante Veränderungen zwischen T2 und T3 in erwarteter Richtung fanden sich für die Erschöpfungsneigung, die Frustration am Ende der Schicht und der Vitalitätsveränderung über die Schichtmessungen.
- Auch bei der Wartegruppe ändert sich die „Hilfsfrist“ nach dem Training nicht.
- Im langfristigen Verlauf werden die geistigen Anforderungen geringer wahrgenommen; wahrgenommene zeitliche Anforderungen, Frustration, mittlere und minimale Herzfrequenz sinken mit mittleren Effektstärken, ohne dass diese Veränderung signifikant wird.
- Die Hinzunahme der Dropout-Daten verändert die Ergebnismuster der gepaarten Stichproben auf der Resultateebene nicht systematisch.

## 5. Moderationseffekte von Erfahrung

Zusätzlich zu den Trainingseffekten, die die Hauptfragestellung der vorliegenden Studie darstellen, wird in den folgenden Analysen Erfahrung als zentraler Aspekt analysiert, für den auch im Hinblick auf das Training ein moderierender Einfluss angenommen wurde. Abweichend von dem üblichen Vorgehen für Moderationsanalysen sensu Baron und Kenny (1986), das voraussetzen würde, dass für eine Vielzahl von Erfahrungsvariablen Analysen durchgeführt werden, wird im Folgenden ein regressionsanalytisches Vorgehen mit Veränderungswerten gewählt. Da Erfahrung häufig eng mit dem Alter einer Person korreliert ist, stellt sich hier allerdings ein Multikollinearitätsproblem, so dass jeweils der Variance Inflation Factor (VIF) und die Toleranz beachtet wurden. Es werden entsprechend nur Erfahrungsvariablen aufgenommen, die nicht zu einer übermäßigen Varianzinflationierung führen (Faustregel  $\text{Max VIF} \approx 5$ ) und für die ausreichend Werte vorliegen.

Betrachtet werden exemplarisch diejenigen Variablen, in denen sich relativ belastbare Veränderungen gezeigt haben. Dazu werden jeweils die Werte von T2 (kurzfristig) T4 (langfristig) als abhängige Variable in den Analysen verwendet und jeweils nur die Gruppen genutzt, die im entsprechenden Zeitraum das Training absolviert haben, also für die kurzfristigen Berechnungen die Gruppen 1 und 3, für die langfristigen alle drei Gruppen.

Im ersten Schritt werden die demographischen Daten Alter und Geschlecht aufgenommen, im zweiten jeweils der Testwerte der abhängigen Variablen zum ersten Messzeitpunkt. Im dritten Schritt werden die Erfahrungsvariablen aufgenommen um zu sehen, ob die Erfahrung zusätzlich einen signifikanten Beitrag zur Änderung der abhängigen Variablen liefern kann. Als Erfahrungsvariablen werden die Erfahrung im Rettungsdienst (Qualifikation und Monate im Rettungsdienst) sowie im Straßenverkehr (durchschnittliche Fahrstunden als Fahrer im Rettungsdienst pro Monat, Anzahl der Führerscheine sowie die Fahrerfahrung basierend auf den Jahren des Führerscheinbesitzes) herangezogen. Die Jahreskilometerfahrleistung fehlt in der Stichprobe zu oft, um sie als Prädiktor aufnehmen zu können.

### 5.1 Einfluss der Erfahrung auf kurzfristige Trainingseffekte

Belastbare kurzfristige Trainingseffekte konnten nur im Wissenstest, der Risikowahrnehmung bei normalen Fahrten sowie der maximalen und durchschnittlichen Geschwindigkeit gefunden werden. Diese Variablen werden explizit hinsichtlich des Einflusses der Erfahrung dargestellt (vgl. Tabelle 45 und Tabelle 46).

Die vollständigen Modelle sind jeweils bei allen vier abhängigen Variablen ausschließlich durch den Ausgangswert der jeweiligen Variable signifikant, die Varianzaufklärung durch die Erfahrungsvariablen jenseits von Alter, Geschlecht und Ausgangswert liegen mit knapp 2% beim Wissen bis 5% für die Risikowahrnehmung und maximale Geschwindigkeit in geringen Bereichen. Auch die einzelnen Aspekte von Erfahrung haben kaum prädiktiven Wert. Eine Ausnahme stellt hierbei die Qualifikation im Rettungsdienst mit ihrem Einfluss auf die Risikowahrnehmung bei Normalfahrten sowie der maximalen Geschwindigkeit dar. Beide Male zeigen die  $\beta$  einen Zusammenhang in die Richtung, dass höher qualifizierte Teilnehmer auch bei kontrolliertem Ausgangswert eine stärkere Zunahme der Risikowahrnehmung und eine höhere Abnahme der Maximalgeschwindigkeit durch das Training zeigen als niedriger qualifizierte Teilnehmer. Für die Erfahrungsaspekte aus dem Bereich des Fahrens zeigen sich keinerlei signifikanten Einflüsse, allerdings fällt auf, dass hier häufige Vorzeichenwechsel daraufhin deuten, dass diese Variablen auf unterschiedliche abhängige Variablen differenziell einwirken.

Tabelle 45: Regression zur Kontrolle des Erfahrungseinflusses auf die kurzfristigen Effekte des Trainings (I)

Abhängige Variable	Post-Wissenstest (N=93)			T2 Risikowahrnehmung normaler Fahrten (N=89)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.048	.048		.060 <sup>+</sup>	.060 <sup>+</sup>
...Alter	-0.18 <sup>+</sup>			0.24*		
...Geschlecht	-0.07			-0.15		
Schritt 2: T1 Messwert		.168***	.120***		.463***	.402***
...Ausgangswert	0.35***			0.66***		
Schritt 3: Erfahrung		.185*	.018		.512***	.050
...Qualifikation	-0.02			0.18*		
...RD-Monate	0.20			-0.09		
...Fahrleistung RD/Monat	0.04			0.12		
...Anzahl Führerschein	-0.08			0.01		
...Fahrerfahrung	-0.09			0.17		

Berechnung: hierarchische Regression über die Differenz der abhängigen Variablen zwischen den Zeitpunkten T1 und T2; Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Tabelle 46: Regression zur Kontrolle des Erfahrungseinflusses auf die kurzfristigen Effekte des Trainings (II)

Abhängige Variable	T2 Mittlere Geschwindigkeit (N=56)			T2 Maximale Geschwindigkeit (N=56)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.001	.001		.031	.031
...Alter	0.01			-0.09		
...Geschlecht	0.03			-0.14		
Schritt 2: T1 Messwert		.634***	.633***		.613***	.582***
...Ausgangswert	0.80***			0.77***		
Schritt 3: Erfahrung		.662***	.028		.664***	.051
...Qualifikation	0.06			-0.19*		
...RD-Monate	-0.16			0.14		
...Fahrleistung RD/Monat	-0.01			-0.09		
...Anzahl Führerschein	0.14			-0.06		
...Fahrerfahrung	-0.08			0.10		

Berechnung: schrittweise Regression über die Differenz der abhängigen Variablen zwischen den Zeitpunkten T1 und T2; Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001  
 Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

## 5.2 Einfluss der Erfahrung auf langfristige Trainingseffekte

Langfristige Trainingseffekte wurden insgesamt für die Unverträglichkeit, die mittlere und maximale gefahrene Geschwindigkeit, die mittlere positive Längsbeschleunigung und die maximale Bremsverzögerung sowie für die wahrgenommenen geistigen Anforderungen zum Schichtende gefunden. Tabelle 47 bis Tabelle 49 beschreiben die Einflüsse der Erfahrungsvariablen für diese abhängigen Variablen.

Tabelle 47: Regression zur Kontrolle des Erfahrungseinflusses für die langfristigen Effekte des Trainings (I)

Abhängige Variable	T4 Unverträglichkeit (N=43)			T4 Geistige Anforderungen (N=36)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.143*	.143*		.149 <sup>+</sup>	.149 <sup>+</sup>
...Alter	-0.36*			0.41*		
...Geschlecht	-0.04			-0.20		
Schritt 2: T1 Messwert		.497***	.354***		.546***	.397***
...Ausgangswert	0.60***			0.65***		
Schritt 3: Erfahrung		.539***	.042		.592***	.045
...Qualifikation	0.12			-0.14		
...RD-Monate	-0.07			0.01		
...Fahrleistung RD/Monat	0.12			0.03		
...Anzahl Führerschein	-0.03			0.18		
...Fahrerfahrung	0.16			0.00		

Berechnung: schrittweise Regression über die Differenz der abhängigen Variablen zwischen den Zeitpunkten T1 und T4; Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Ähnlich wie für die kurzfristigen Trainingseffekte finden sich bei den Regressionsanalysen für die längerfristigen Effekte nur selten Zusammenhänge, die über einen Einfluss des Ausgangswerts der Messung hinausgehen, und auch der Einfluss des Prädiktors „Ausgangswert“ wird über die Zeit eher kleiner, v.a. für die Fahrprofilaten. Einflüsse der Kontrollvariablen mit signifikanten Varianzaufklärungen um die 15% finden sich für die Unverträglichkeit und die wahrgenommenen geistigen Anforderungen. Wesentlicher Prädiktor ist das Alter – jüngere Teilnehmer haben höhere Unverträglichkeitswerte und ältere geben höhere geistige Anforderungen zu T4 an.

Tabelle 48: Regression zur Kontrolle des Erfahrungseinflusses für die langfristigen Effekte des Trainings (II)

Abhängige Variable	T4 Mittlere Geschwindigkeit (N=29)			T4 Maximale Geschwindigkeit (N=29)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.019	.019		.023	.023
...Alter	-0.06			0.03		
...Geschlecht	0.14			0.14*		
Schritt 2: T1 Messwert		.229 <sup>+</sup>	.210*		.363**	.341***
...Ausgangswert	0.56*			0.77***		
Schritt 3: Erfahrung		.381	.151		.565*	.202
...Qualifikation	-0.23			-0.02		
...RD-Monate	0.54			0.76*		
...Fahrleistung RD/Monat	-0.10			-0.09		
...Anzahl Führerschein	-0.14			-0.22		
...Fahrerfahrung	0.25			-0.10		

Berechnung: schrittweise Regression über die Differenz der abhängigen Variablen zwischen den Zeitpunkten T1 und T4; Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001  
 Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

Für die verschiedenen Aspekte der Erfahrung zeigen sich zwar mit zusätzlichen Varianzaufklärungen zwischen 4 % (Unverträglichkeit) und 20% (maximale Geschwindigkeiten) zum Teil recht hohe, aber in keinem Fall signifikante Zuwächse. Ähnlich wie bei den Analysen zur kurzfristigen Prädiktion sieht es bei den standardisierten Regressionskoeffizienten im langfristigen Vergleich aus: Ein einzelner Erfahrungsaspekt, die Anzahl der Monate im Rettungsdienst – also die Berufserfahrung – liefert hohe und z.T. signifikante  $\beta$ . Die maximale Geschwindigkeit zu T4, aber auch die mittlere Geschwindigkeit und mittlere Längsbeschleunigung sind über den Einfluss des Ausgangswerts dieser Maße hinaus bei Personen mit langer Berufserfahrung höher als bei Personen mit weniger Berufserfahrung.

Tabelle 49: Regression zur Kontrolle des Erfahrungseinflusses für die langfristigen Effekte des Trainings (III)

Abhängige Variable	T4 Mittlere Längsbeschleunigung (N=29)			T4 Maximale Bremsverzögerung (N=29)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.038	.038		.028	.028
...Alter	-0.16			0.17		
...Geschlecht	0.15			-0.01		
Schritt 2: T1 Messwert		.051	.013		.271*	.243**
...Ausgangswert	0.12			0.53**		
Schritt 3: Erfahrung		.210	.159		.391	.120
...Qualifikation	0.26			-0.26		
...RD-Monate	0.56			0.15		
...Fahrleistung RD/Monat	-0.09			0.22		
...Anzahl Führerschein	0.07			0.18		
...Fahrerfahrung	-0.26			0.43		

Berechnung: schrittweise Regression über die Differenz der abhängigen Variablen zwischen den Zeitpunkten T1 und T4; Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001  
 Fahrprofilaten enthalten lediglich RTW-Fahrten zum Einsatzort

Für alle anderen Erfahrungsaspekte, insbesondere erneut für die Aspekte der Fahrerfahrung, zeigen sich keine signifikanten  $\beta$ , dafür aber erheblich Wechsel in den Vorzeichen, die wieder auf differenzielle Zusammenhänge hinweisen. Damit lässt sich abschließend für den Einfluss von Erfahrung in den trainierten Gruppen nur konstatieren, dass sie kaum prädiktiv für die erfassten Trainingsevaluationsmaße sind und dort, wo sich Effekte finden, diese eher aus dem Bereich der beruflichen Erfahrung stammen als aus dem Bereich des Fahrens.

## 6. Erfolgsfaktoren

Die bisherigen Auswertungen erfolgten in der Logik klassischer Evaluationen, ohne zu berücksichtigen, dass es jenseits von Gruppeneffekten individuelle Trainingserfolge geben kann, die im Sinne der „klinischen“ Bedeutsamkeit eine Rolle spielen können. Kazdin (1999) etwa definiert klinische Signifikanz oder Bedeutsamkeit als bezogen auf „the practical or applied value or importance of the effect of the intervention - that is, whether the intervention makes a real (e.g., genuine, palpable, practical, noticeable) difference in everyday life to the clients or to others with whom the client interacts“ (S. 332). Um diese Form der Bedeutsamkeit bei der Bewertung von Effekten beurteilen zu können, sind inhaltliche Kriterien anzulegen. Bislang mangelt es jedoch an etablierten Konventionen zur Interpretation solcher Unterschiede bei verschiedenen Antwortformaten. Mit einer gewissen Willkür werden bisweilen prozentuale Abweichungen zwischen Messwerten vor und nach der Intervention (z. B. 10 %) als mögliches Kriterium genannt (vgl. auch Herbig & Glaser, 2013). Für die folgenden Auswertungen wird ein anderes Vorgehen gewählt, das in Abschnitt 6.1 näher beschrieben wird. Hintergrund der Auswertungen ist der Vergleich erfolgreicher und nicht erfolgreicher Teilnehmer hinsichtlich verschiedener Aspekte der individuellen, organisationalen und Trainingsgruppenebene mit dem Ziel, praktische Hinweise zu den folgenden Fragen zu erhalten: Welche Voraussetzungen sollte ein Teilnehmer mitbringen, um vom Training zu profitieren? Wie sollten Trainingsgruppen zusammengesetzt werden, damit die einzelnen Teilnehmer erfolgreich sind? Und, welche Organisationen profitieren bzw. was können Organisationen tun, um den Trainingserfolg zu fördern? Anders als in den varianzanalytischen Auswertungen, in denen ein Ausschluss aufgrund unterschiedlicher Datenniveaus notwendig war, werden bei dieser Betrachtung auch die Verhaltens-/Fahrprofilaten von Teilnehmern betrachtet, die im Verlauf der Studie nur NEF gefahren sind.

### 6.1 Bildung von Indices zum allgemeinen und ebenen-spezifischen individuellen Trainingserfolg

Da es sich bei den untersuchten abhängigen Variablen um nur schwer änderbare Einstellungs- und – im Falle des Fahrstils auch für andere Personen höchstrelevante - Verhaltensmerkmale handelt bei einer vergleichsweise geringen „Dosierung“ eines eintägigen Trainings wird der individuelle Trainingserfolg bereits bei kleinen positiven Entwicklungen in die gewünschte Richtung gesehen. Daher werden zunächst für alle interessierenden Werte individuelle Differenzen ( $\Delta$ ) gebildet, und zwar für den kurzfristigen Erfolg jeweils vor und kurz nach dem Training (Interventionsgruppen 1 und 3 Messzeitpunkte T1/T2, Wartegruppe 2 Messzeitpunkte T2/T3) und für den längerfristigen Erfolg für alle Gruppen die Differenz zwischen T1 und T4. Ein Trainingserfolg in der entsprechenden Variable wird dann festgehalten, wenn der individuelle Wert überzufällig ansteigt/sinkt. Genauer wird dies wie folgt definiert:

Für  $\Delta = 0 \pm SE_i$  des jeweiligen Werts, Erfolg  $E_i=0$ ;

für  $\Delta < 0 \pm SE_i$ ,  $E_i=-1$  und

für  $\Delta > 0 \pm SE_i$ ,  $E_i=1$

wobei Standardfehler (SE) die Zufallsfluktuation;  $E_i=0$  keine oder zufällige Veränderungen,  $E_i=-1$  Veränderungen in unerwünschte Richtungen und  $E_i=1$  Veränderung in die gewünschte Richtung (Erfolg) darstellt. Ergänzend zu dieser Definition mit einer sehr geringen Schwelle für „Erfolg“ werden für die verschiedenen Ebenen Trainingserfolge wie folgt als Summenwerte der einzelnen Merkmale gebildet:

$$\sum_{i=1}^n E_i$$



wobei Werte  $> 0$  den Ebenenerfolg angeben und Werte  $\leq 0$  als nicht erfolgreich betrachtet werden. Durch diese Form der Indexbildung können Teilnehmer mit Verbesserungen in einzelnen Variablen Verschlechterungen in anderen Variablen ausgleichen, müssen aber mindestens Verbesserungen in einer Variablen (mehr) als Variablen mit Verschlechterungen aufweisen. Zu beachten ist hier auch, dass Ergebnisse des Wissenstest (vgl. Abschnitt 4.2.2.1) ausschließlich als Einzelparameter betrachtet werden können und nicht in den Ebenenindex „Lernerfolg“ einfließen, da die Wartegruppe keinen Wissenstest nach dem Training erhielt und es damit zu einer systematischen Verzerrung kommen könnte.

Da nicht alle Einzelindikatoren eine vergleichbar hohe „Wertigkeit“ für die Verkehrssicherheit als zentralem Ziel des Trainings aufweisen und daher auch ein Ausgleich von Verschlechterungen über verschiedene Ebenen hinweg nicht wünschenswert ist, wird zudem ein weiterer allgemeiner Erfolgsindex über alle Ebenen und Merkmale eingefügt, der sich aus der Summe des gewichteten Erfolgs verschiedener als besonders relevant erachteter Variablen zusammensetzt, während als nicht so wesentlich eingeschätzte Aspekte nicht in diese Bewertung einfließen. Die höchsten Gewichtungen werden auf der Verhaltensebene gesetzt, hier gehen mit dreifacher Gewichtung jeweils die durchschnittliche und maximale gefahrene Geschwindigkeit ein sowie mit zweifacher Gewichtung die jeweils mittlere Beschleunigung und mittlere Bremsung. Zusätzlich werden mit 1.5-facher Wertung die Anzahl von Extremwerten (Trigger) bei Beschleunigung und Bremsung aufgenommen. Damit können auf der Verhaltensebene maximal 13 Punkte erreicht werden. An zweiter Stelle in der Bedeutsamkeit der Indikatoren werden Aspekte der Lernebene gewertet, wobei positive Entwicklungen in den verschiedenen Aspekten der Risikowahrnehmung mit 1.5-facher Gewichtung einfließen und positive Veränderungen in den verkehrsbezogenen Persönlichkeitsmerkmalen mit einfacher Wertung, so dass insgesamt für die Lernebene ein maximaler Punktwert von 11 möglich ist. Als der Bereich, der am ehesten von nicht durch das Projekt beeinflussbaren Bedingungen abhängt, erhält die Resultateebene mit insgesamt 9 Punkten das geringste Gewicht in dem Indikator des Gesamterfolgs. Dabei werden Irritation, Wohlbefinden, psychosomatische Beschwerden und Burnoutaspekte jeweils mit 0.5-fach gewertet, die schichtbezogenen Aspekte des emotionalen Befindens, die wahrgenommenen zeitlichen und geistigen Anforderungen und empfundene Anstrengung sowie die Ermüdung über die Schicht einfach und die mittlere Herzfrequenz 1.5-fach.

Um zu große Verzerrungen durch fehlende Werte zu vermeiden, wird der Gesamterfolg nur für diejenigen Teilnehmer berechnet, die aufgrund ihrer vorhandenen Veränderungs-messungen mindestens 17 dieser 33 möglichen Punkte hätten erreichen können. Schließlich werden diejenigen Personen als insgesamt erfolgreiche Studienteilnehmer angesehen, die mehr als 50% der für sie individuell möglichen Maximalpunktzahl erreicht haben.

## 6.2 Anzahl erfolgreicher Teilnehmer

Tabelle 50 gibt einen Überblick über die Anzahl von Teilnehmern, die sich in einzelnen untersuchten Merkmalen gemäß der dargestellten Definition kurzfristig oder längerfristig verbessert oder verschlechtert haben. Insgesamt zeigt sich ein sehr durchmischtes Bild, insbesondere im Hinblick darauf, dass es in jedem erfassten Merkmal (mit Ausnahme der Trigger-Anzahl für extreme Querbeschleunigung) neben verbesserten und unveränderten Werten auch verschlechterte Werte gibt. Bereiche, in denen sich die Hälfte oder mehr aller Trainingsteilnehmer (wenn auch u.U. nur leicht) kurz nach dem Training verbesserten, sind alle Parameter der Risikowahrnehmung, die mittlere und maximale Geschwindigkeiten bei Sondersignalfahrten sowie die maximale Beschleunigung bei diesen Fahrten, Irritation und körperliche Beschwerden sowie die während der Schichten gemessene Ermüdung und Veränderung in der Vigilanz und die wahrgenommenen geistigen und zeitlichen Anforderungen. Im Hinblick auf längerfristige Veränderungen bleiben die Tendenzen größtenteils bestehen, oft ist, wie zu erwarten, der Anteil der Personen mit verbesserten Werten jedoch kleiner als bei den Daten zum kurzfristigen Erfolg - besonders deutlich ist dies bei

Geschwindigkeit und Beschleunigung auf der Verhaltensebene, aber auch bei der Risikowahrnehmung auf der Lernebene. Anderes sieht es hingegen bei den bereits von Beginn an längerfristig konzeptualisierten Resultaten; hier liegen - vor allem bei der mittleren Herzfrequenz - die prozentualen Werte für Personen mit Verbesserungen sogar über denen des kurzfristigen Erfolgs.

Tabelle 50: Prozentuale Verteilung des Trainingserfolgs in einzelnen Merkmalen

	Kurzfristiger Erfolg (Anzahl/Prozent)			Längerfristiger Erfolg (Anzahl/Prozent)		
	verbessert	unverändert	verschlechtert	verbessert	unverändert	verschlechtert
<b>Lernebene</b>						
Wissen*	67 / 69.8	6 / 6.3	23 / 24.0	/	/	/
Extraversion	62 / 40.3	37 / 24.0	55 / 35.7	21 / 46.7	6 / 13.3	18 / 40.0
Sensation Seeking	65 / 42.2	42 / 27.3	47 / 30.5	17 / 37.8	14 / 31.1	14 / 31.1
Gewissenhaftigkeit	52 / 34.0	50 / 32.7	51 / 33.3	15 / 34.1	12 / 27.3	17 / 38.6
Unverträglichkeit	69 / 44.8	37 / 24.0	48 / 31.2	<b>23 / 51.1</b>	10 / 22.2	12 / 26.7
Reaktanz	62 / 40.3	36 / 23.4	56 / 36.4	12 / 26.7	16 / 35.6	17 / 37.8
Risikowahrnehmung bei schweren Regelverstößen	<b>77 / 50.0</b>	23 / 14.9	54 / 35.1	22 / 48.9	10 / 22.2	13 / 28.9
Risikowahrnehmung bei mittelschweren Regelverstößen	<b>85 / 55.2</b>	28 / 18.2	41 / 26.6	<b>27 / 60.0</b>	5 / 11.1	13 / 28.9
Risikowahrnehmung bei leichten Regelverstößen	<b>91 / 59.1</b>	19 / 12.3	44 / 28.6	21 / 46.7	4 / 8.9	20 / 44.4
Risikowahrnehmung bei normalem Fahren	<b>90 / 58.4</b>	26 / 16.9	38 / 24.7	22 / 48.9	8 / 17.8	15 / 33.3
Risk Seeking	58 / 37.9	46 / 30.1	49 / 32.0	21 / 46.7	8 / 17.8	16 / 35.6
<b>Verhaltensebene</b>						
Durchschnittliche Geschwindigkeit (in km/h)	<b>67 / 55.8</b>	9 / 7.5	44 / 36.7	14 / 46.7	8 / 26.7	8 / 26.7
Maximale Geschwindigkeit (in km/h)	<b>69 / 57.5</b>	10 / 8.3	41 / 34.2	11 / 36.7	9 / 30.0	10 / 33.3
Maximale positive Längsbeschleunigung (in m/s <sup>2</sup> )	<b>63 / 52.5</b>	13 / 10.8	44 / 36.7	10 / 33.3	11 / 36.7	9 / 30.0
Mittlere positive Längsbeschleunigung (in m/s <sup>2</sup> )	56 / 46.7	14 / 11.7	50 / 41.7	<b>19 / 63.3</b>	7 / 23.3	4 / 13.3
Anzahl Trigger pro Minute positive Längsbeschleunigung (5,9 m/s <sup>2</sup> )	8 / 6.7	108 / 90.0	4 / 3.3	5 / 16.7	25 / 83.3	0 / 0.0
Maximale negative Längsbeschleunigung (in m/s <sup>2</sup> )	43 / 35.8	19 / 15.8	58 / 48.3	9 / 30.0	8 / 26.7	13 / 43.3
Mittlere negative Längsbeschleunigung (in m/s <sup>2</sup> )	53 / 44.2	13 / 10.8	54 / 45.0	3 / 10.0	10 / 33.3	17 / 56.7
Anzahl Trigger pro Minute negative Längsbeschleunigung (3 m/s <sup>2</sup> )	14 / 11.7	94 / 78.3	12 / 10.0	7 / 23.3	23 / 76.7	0 / 0.0
Maximale positive Querschleunigung (in m/s <sup>2</sup> )	58 / 48.3	22 / 18.3	40 / 33.3	14 / 46.7	11 / 36.7	5 / 16.7
Mittlere positive Quer-	58 / 48.3	20 / 16.7	42 / 35.0	<b>21 / 70.0</b>	6 / 20.0	3 / 10.0

Abschlussbericht „Achtung Blaulicht“

	Kurzfristiger Erfolg (Anzahl/Prozent)			Längerfristiger Erfolg (Anzahl/Prozent)		
	verbessert	unverändert	verschlechtert	verbessert	unverändert	verschlechtert
beschleunigung (in m/s <sup>2</sup> )						
Maximale negative Querschleunigung (in m/s <sup>2</sup> )	40 / 33.3	21 / 17.5	59 / 49.2	5 / 16.7	14 / 46.7	11 / 36.7
Mittlere negative Querschleunigung (in m/s <sup>2</sup> )	43 / 35.8	24 / 20.0	53 / 44.2	2 / 6.7	10 / 33.3	18 / 60.0
Anzahl Trigger pro Minute Querschleunigung (6,9 m/s <sup>2</sup> )	6 / 5.0	114 / 95.0	0 / 0.0	6 / 20.0	24 / 80.0	0 / 0.0
<b>Resultateebene</b>						
Irritation gesamt	<b>87 / 56.5</b>	16 / 10.4	51 / 33.1	<b>27 / 60.0</b>	10 / 22.2	8 / 17.8
GBB Beschwerdedruck	<b>81 / 55.9</b>	15 / 10.3	49 / 33.8	<b>25 / 58.1</b>	5 / 11.6	13 / 30.2
Wohlbefinden (WHO-5)	63 / 41.2	23 / 15.0	67 / 43.8	19 / 43.2	4 / 9.1	21 / 47.7
Burnout (emotionale Erschöpfung)	60 / 39.2	25 / 16.3	68 / 44.4	13 / 29.5	12 / 27.3	19 / 43.2
Burnout (Zynismus)	54 / 35.3	39 / 25.5	60 / 39.2	<b>22 / 50.0</b>	9 / 20.5	13 / 29.5
Vitalität (BBS)	66 / 48.9	11 / 8.1	58 / 43.0	13 / 38.2	7 / 20.6	14 / 41.2
Gleichgewicht (BBS)	57 / 42.2	15 / 11.1	63 / 46.7	9 / 26.5	8 / 23.5	17 / 50.0
Extraversion (BBS)	52 / 38.5	16 / 11.9	67 / 49.6	10 / 29.4	3 / 8.8	21 / 61.8
Vigilanz (BBS)	<b>71 / 52.6</b>	11 / 8.1	53 / 39.3	12 / 35.3	8 / 23.5	14 / 41.2
NASA geistige Anforderungen	<b>73 / 55.3</b>	9 / 6.8	50 / 37.9	<b>22 / 59.5</b>	6 / 16.2	9 / 24.3
NASA zeitliche Anforderungen	<b>69 / 52.3</b>	12 / 9.1	51 / 38.6	<b>23 / 62.2</b>	2 / 5.4	12 / 32.4
NASA Anstrengung	63 / 47.7	11 / 8.3	58 / 43.9	17 / 45.9	4 / 10.8	16 / 43.2
Mittlere Ermüdung (KSS)	<b>65 / 50.8</b>	7 / 5.5	56 / 43.8	<b>22 / 59.5</b>	3 / 8.1	12 / 32.4
Mittlere Herzfrequenz (Schläge pro Minute)	46 / 47.9	3 / 3.1	47 / 49.0	<b>12 / 66.7</b>	3 / 16.7	3 / 16.7
Maximale Herzfrequenz (Schläge pro Minute)	42 / 43.8	14 / 14.6	40 / 41.7	6 / 33.3	6 / 33.3	6 / 33.3

Fett hervorgehobene Werte – 50% und mehr Personen mit verbesserten Werten; \* nur Gruppen 1 und 3

Erlaubt man, wie in den ebenenspezifischen Erfolgsindices, dass negative Entwicklungen durch positive ausgeglichen werden, haben jeweils mehr als die Hälfte der Trainingsteilnehmer von den Trainings profitiert und dies sowohl kurz- wie längerfristig (vgl. Tabelle 51). Der größte Anteil erfolgreicher Teilnehmer findet sich kurzfristig für die Lernebene gefolgt von der Resultateebene, während längerfristig die Resultateebene den höchsten Anteil erfolgreicher Teilnehmer aufweist und die Anteile auf der Lern- und Verhaltensebene sinken. Interessant ist hierbei, dass auf der Lernebene etwa 79% der Teilnehmer, die unmittelbar von dem Training profitierten, diese positiven Veränderungen auch später noch zeigten. Dieser Anteil ist mit etwa 56% im Bereich des Verhaltens deutlich geringer. Dies verweist darauf, dass bei einigen Personen Lernerfolge mit der Zeit wieder verblassen, es gleichzeitig aber auch Personen gibt, die erst nach einem längeren Zeitraum nach dem Training erfolgreiche Veränderungen zeigen.

Für das gewichtete globale Erfolgsmaß ist der Anteil der erfolgreichen Teilnehmer naturgemäß geringer: knapp 38% der Teilnehmer können kurz nach dem Training als insgesamt erfolgreich klassifiziert werden, 37% sind dies längerfristig (vgl. Tabelle 51).

Tabelle 51: Prozentuale Verteilung des Trainingserfolgs auf den verschiedenen Evaluationsebenen und global

	Kurzfristiger Erfolg (Anzahl/Prozent)		Längerfristiger Erfolg (Anzahl/Prozent)		
	erfolgreich	nicht erfolgreich	erfolgreich	nicht erfolgreich	nachhaltig erfolgreich*
<b>Ebenenspezifischer Erfolg</b>					
Lernen	91 / 59.1	63 / 40.9	24 / 53.3	21 / 46.7	19 / 79.2
Verhalten	65 / 54.2	55 / 45.8	16 / 53.3	14 / 46.7	9 / 56.3
Resultate	92 / 56.1	72 / 43.9	29 / 58.0	21 / 42.0	18 / 62.1
<b>Globaler Erfolg</b>					
Gesamttrainingserfolg	50 / 37.9	82 / 62.1	13 / 37.1	22 / 62.9	8 / 61.5

\*Nachhaltig erfolgreich = Personen, die längerfristigen Erfolg haben und bereits kurz nach dem Training erfolgreiche Veränderungen zeigten (Absolut / Prozent aller Personen mit längerfristigem Erfolg)

In den folgenden Abschnitten werden diese Trainingserfolgsvariablen genutzt, um festzustellen, ob es individuelle (vgl. Abschnitt 6.3), organisationsbezogene (vgl. Abschnitt 6.4) oder trainingsgruppenbezogene (vgl. Abschnitt 6.5) Aspekte gibt, die zwischen erfolgreichen und nichterfolgreichen Teilnehmern auf den verschiedenen Ebenen und global unterscheiden. Der verwendete Begriff „Prädiktion“ ist dabei nicht statistisch zu verstehen, dass es um die retrospektive Identifikation von Voraussetzungen geht, die dazu führen könnten, dass das Training zu positiven Veränderungen führt. Ziel dieser Auswertung ist, erste Hinweise zu erhalten, ob es spezifische Gruppen von Personen gibt, für die das Training erfolgsversprechend ist und ob ggf. bei der Zusammensetzung der Trainingsteilnehmer und der Arbeitsorganisation in den „Herkunftsorganisationen“ der Teilnehmer auf bestimmte Aspekte zu achten ist, um die Erfolgswahrscheinlichkeit zu erhöhen.

### 6.3 Prädiktion von Erfolg durch individuelle Voraussetzungen

Im folgenden Abschnitt werden vor dem Hintergrund der dargestellten Fragestellung zunächst individuelle Voraussetzungen der Trainingsteilnehmer betrachtet. Hierzu werden in Tabelle 52 bis Tabelle 55 jeweils Merkmale der Trainingsteilnehmer, wie sie zum ersten Messzeitpunkt erhoben wurden, in ihrer Verteilung auf kurz- oder längerfristige erfolgreiche und nichterfolgreiche Personen auf den verschiedenen Evaluationsebenen dargestellt.

Insgesamt finden sich nur wenige individuelle Voraussetzungen, die systematisch signifikant zwischen erfolgreichen und nichterfolgreichen Trainingsteilnehmern unterscheiden. Im Bereich des kurzfristigen Erfolgs auf der Lernebene (vgl. Tabelle 52) finden sich tendenzielle Unterschiede hinsichtlich Geschlecht und sehr signifikante Unterschiede hinsichtlich dem Spaß, den die Teilnehmer am Fahren mit Sondersignal haben.

Tabelle 52: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Lernebene durch individuelle Voraussetzungen

Merkmal	Kurzfristig erfolgreich: Lernebene, n ≤ 91	Kurzfristig nicht erfolgreich: Lernebene, n ≤ 63	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Lernebene, n ≤ 24	längerfristig nicht erfolgreich: Lernebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Geschlecht			3.35	p=.082 <sup>+</sup>			5.74	.p=.029*
Männer	65	53			22	13		
Frauen	26	10			2	8		
Alter	M = 33.6, SD = 9.02	M = 32.6, SD = 10.2	-1.04	p=.299 n.s.	M = 37.5, SD = 11.4	M = 32.1, SD = 8.94	-1.57	p=.118 n.s.
Qualifikation			2.55	p=.517 n.s.			0.97	p=1.00 n.s.
Rettungsdiensthelfer	2	2			1	0		
Rettungssanitäter	18	17			4	3		
Rettungsassistent	69	44			19	18		
Notfallsanitäter	2	0			0	0		
Dauer Beschäftigung im Rettungsdienst (Monate)	M = 130.7 SD = 98.82	M = 126.1 SD = 114.39	-0.70	p=.486 n.s.	M = 178.8 SD = 127.9	M = 115.5 SD = 90.32	-1.78	p=.076 <sup>+</sup>
Beschäftigungsart			2.17	p=.268 n.s.			2.24	p=.501 n.s.
hauptberuflich	84	54			21	19		
nebenberuflich	2	4			1	2		
ehrenamtlich	5	5			2	0		
Bundesland			1.43	p=.305 n.s.			0.04	p=1.00 n.s.
Bayern	55	44			13	12		
Sachsen	36	19			11	9		
Anzahl Führerscheinklassen	M = 4.22, SD = 1.63	M = 4.48, SD = 1.61	-1.03	p=.303 n.s.	M = 4.46, SD = 1.44	M = 4.24, SD = 1.61	-0.60	p=.557 n.s.
Jahre Führerscheinbesitz	M = 15.0; SD = 8.81	M = 13.9; SD = 9.48	-1.12	p=.262 n.s.	M = 18.7; SD = 10.9	M = 13.0; SD = 8.08	-1.67	p=.096 <sup>+</sup>
Erlaubnis zur Fahrgastbeförderung			0.33	p=.682 n.s.			0.16	p=.746 n.s.
ja	17	14			7	5		
nein	72	47			17	16		
Jährliche Fahrleistung gesamt in km (alle Klassen)	M = 30441 SD = 15158.9	M = 34199 SD = 15719.2	-1.38	p=.169 n.s.	M = 29855 SD = 12306.3	M = 36653 SD = 14806.2	-1.73	p=.084 <sup>+</sup>
Punkte Fahreignungsregister			1.59	p=.249 n.s.			3.57	p=.073 <sup>+</sup>
ja	47	39			18	10		

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich: Lernebene, n ≤ 91	Kurzfristig nicht erfolgreich: Lernebene, n ≤ 63	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Lernebene, n ≤ 24	längerfristig nicht erfolgreich: Lernebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
nein	44	24			6	11		
Jemals in Unfall mit Rettungswagen involviert?			0.12	p=.745 n.s.			0.02	p=1.00 n.s.
ja	48	35			15	13		
nein	43	28			9	8		
Anzahl der Unfälle mit RTW als Fahrer	M = 0.57, SD = 0.78	M = 0.60, SD = 0.68	-0.67	p=.509 n.s.	M = 0.63, SD = 0.50	M = 0.71, SD = 0.78	-0.32	p=.791 n.s.
Stunden pro Monat als Fahrer auf RTW	M = 85.9 SD = 51.58	M = 79.9 SD = 51.19	-0.59	p=.557 n.s.	M = 72.9 SD = 52.17	M = 94.89 SD = 52.58	-1.45	p=.150 n.s.
„Freude am Fahren mit Sondersignal“ (Fahrer)	M = 3.47 SD = 0.91	M = 3.88 SD = 1.01	-2.59	p=.009**	M = 3.35 SD = 0.88	M = 3.45 SD = 1.30	-0.43	p=.677 n.s.
Beanspruchung durch Fahren mit SoSi (Fahrer)	M = 2.95 SD = 1.07	M = 2.83 SD = 1.23	-0.71	p=.482 n.s.	M = 2.96 SD = 1.11	M = 3.07 SD = 1.34	-0.48	p=.638 n.s.
Anzahl verkehrsrelevanter Fortbildungen			5.36	p=.147 n.s.			9.70	p=.021*
Keine	37	17			12	3		
1 - 2	42	38			9	17		
3 - 4	7	7			2	0		
5 und mehr	5	1			1	1		
Blaulichteinweisung innerhalb der letzten fünf Jahre?			0.42	p=.659 n.s.			1.65	p=.259 n.s.
nein	17	9			6	2		
ja	69	49			18	18		
Bereits zu T1 guter Fahrer?			0.60	p=.593 n.s.			3.11	p=.184 n.s.
nein	47	35			10	11		
ja	8	9			6	1		

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveaus: †p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

Tabelle 53: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Verhaltensebene durch individuelle Voraussetzungen

Merkmal	Kurzfristig erfolgreich: Verhaltensebene, n ≤ 65	Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 55	χ <sup>2</sup> oder Z- Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤ 16	längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14	χ <sup>2</sup> oder Z- Statistik	Unter- schied <sup>1</sup>
Geschlecht			0.21	p=.819 n.s.			0.43	p=.642 n.s.
Männer	53	43			14	11		
Frauen	12	12			2	3		
Alter	M = 33.0, SD = 9.10	M = 33.4, SD = 9.78	-0.08	p=.934 n.s.	M = 37.2, SD = 10.8	M = 30.9, SD = 6.62	-1.62	p=.107 n.s.
Qualifikation			2.69	p=.446 n.s.			4.64	p=.072 <sup>+</sup>
Rettungsdiensthelfer	1	3			1	0		
Rettungssanitäter	16	12			1	5		
Rettungsassistent	48	39			14	9		
Notfallsanitäter	0	1			0	0		
Dauer Beschäftigung im Rettungsdienst (Monate)	M = 130.3 SD = 104.47	M = 127.7 SD = 98.49	-0.05	p=.964 n.s.	M = 154.4 SD = 112.08	M = 126.8 SD = 81.10	-0.42	p=.689 n.s.
Beschäftigungsart			2.19	p=.390 n.s.			3.21	p=.209 n.s.
hauptberuflich	56	50			15	12		
nebenberuflich	5	1			0	2		
ehrenamtlich	4	4			1	0		
Bundesland			0.07	p=.847 n.s.			0.74	p=.466 n.s.
Bayern	44	36			9	10		
Sachsen	21	19			7	4		
Anzahl Führerscheinklassen	M = 4.42, SD = 1.69	M = 4.02, SD = 1.65	-1.17	p=.244 n.s.	M = 4.69, SD = 1.82	M = 4.36, SD = 1.39	-0.39	p=.717 n.s.
Jahre Führerscheinbesitz	M = 14.6; SD = 8.59	M = 14.6; SD = 9.31	-0.22	p=.824 n.s.	M = 17.9; SD = 9.61	M = 12.6; SD = 6.54	-1.54	p=.127 n.s.
Erlaubnis zur Fahrgastbeförderung			1.54	p=.261 n.s.			0.05	p=1.00 n.s.
ja	16	9			4	4		
nein	45	45			12	10		
Jährliche Fahrleistung gesamt in km (alle Klassen)	M = 31438 SD = 16456.9	M = 31767 SD = 14470.4	-0.17	p=.870 n.s.	M = 36781 SD = 15797.6	M = 33570 SD = 13365.7	-0.54	p=.601 n.s.
Punkte Fahreignungsregister			0.35	p=.582 n.s.			0.07	p=1.00 n.s.

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich: Verhaltensebene, n ≤ 65	Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 55	χ <sup>2</sup> oder Z- Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤16	längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14	χ <sup>2</sup> oder Z- Statistik	Unter- schied <sup>1</sup>
ja	35	32			11	9		
nein	30	22			5	5		
Jemals in Unfall mit Rettungswagen involviert?			0.77	p=.460 n.s.			1.27	p=.299 n.s.
ja	36	26			7	9		
nein	28	28			9	5		
Anzahl der Unfälle mit RTW als Fahrer	M = 0.56 SD = 0.50	M = 0.48 SD = 0.50	-0.81	p=.435 n.s.	M = 0.44 SD = 0.51	M = 0.71 SD = 0.61	-1.31	p=.223 n.s.
Stunden pro Monat als Fahrer auf RTW	M = 81.3 SD = 47.20	M = 86.8 SD = 55.98	-0.35	p=.732 n.s.	M = 78.4 SD = 41.36	M = 108.1 SD = 66.80	-1.19	p=.246 n.s.
„Freude am Fahren mit Sondersignal“ (Fahrer)	M = 3.64 SD = 0.99	M = 3.66 SD = 0.94	-0.25	p=.803 n.s.	M = 3.69 SD = 1.14	M = 3.08 SD = 1.12	-1.46	p=.161 n.s.
Beanspruchung durch Fahren mit SoSi (Fahrer)	M = 3.08 SD = 1.18	M = 2.76 SD = 0.96	-1.51	p=.130 n.s.	M = 3.13 SD = 1.20	M = 3.00 SD = 1.08	-0.27	p=.820 n.s.
Anzahl verkehrsrelevanter Fortbildungen			8.25	p=.041*			1.74	p=.420 n.s.
Keine	23	20			10	7		
1 - 2	37	21			5	7		
3 - 4	3	11			1	0		
5 und mehr	2	2			0	0		
Blaulichteinweisung innerhalb der letzten fünf Jahre?			0.44	p=.627 n.s.			1.33	p=.330 n.s.
nein	13	8			1	3		
ja	48	41			14	11		
Bereits zu T1 guter Fahrer?			1.45	p=.260 n.s.			1.54	p=.357 n.s.
nein	38	45			13	8		
ja	4	10			2	4		

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: <sup>†</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001



Tendenziell scheinen Männer kurzfristig weniger zu lernen durch das Training als Frauen, und Personen, die weniger Spaß am Fahren mit Sondersignal haben, mehr zu profitieren, als Teilnehmer, denen das Blaulichtfahren Spaß macht. Letzteres ist für den längerfristigen Erfolg auf der Lernebene nicht mehr nachzuweisen, während der Unterschied zwischen Männern und Frauen sich umkehrt und deutlicher wird: Längerfristig profitieren Frauen weniger im Bereich der Einstellungsänderungen als Männer. Um längerfristig erfolgreiche Einstellungsänderungen durch das Training hervorzubringen, scheinen auch weitere Aspekte, v.a. aus dem Bereich der Rettungsdienst- und Fahrerfahrung, eine Rolle zu spielen. So profitieren Personen mit längerer Beschäftigung im Rettungsdienst und längerem Führerscheinbesitz bei gleichzeitig geringerer Fahrleistung pro Jahr und Personen, die bisher an keinen verkehrsrelevanten Fortbildungen teilgenommen hatten. Zudem waren längerfristige Erfolge auf der Lernebene häufiger bei Personen zu finden, die angaben, schon Punkte im Fahreignungsregister zu haben.

Für die individuellen Voraussetzungen, die zwischen verhaltensbezogenen erfolgreichen und nicht-erfolgreichen Teilnehmern unterscheiden (vgl. Tabelle 53), findet sich kurzfristig erneut ein Effekt der Anzahl verkehrsrelevanter Fortbildungen: Ähnlich wie für den längerfristigen Erfolg auf der Lernebene scheinen erfolgreiche Teilnehmer auf der Verhaltensebene dadurch gekennzeichnet zu sein, dass sie vor dem untersuchten Training an weniger Fortbildungen teilgenommen hatten als die nichterfolgreichen Teilnehmer. Weitere Unterschiede für den kurzfristigen Erfolg auf der Verhaltensebene finden sich nicht. Längerfristig sind vor allem höherqualifizierte Rettungsassistenten erfolgreich, während sich in der nicht-erfolgreichen Gruppe proportional vor allem Rettungsassistenten finden. Ohne ein konventionelles Signifikanzniveau zu erreichen, aber mit entsprechenden Tendenzen, die sich auch schon auf der Lernebene zeigten, profitieren ältere Teilnehmer und Personen mit längerem Führerscheinbesitz längerfristig mehr als jüngere und (vermutlich nicht unabhängig vom Alter) Personen mit kürzerem Führerscheinbesitz.

Bei der Betrachtung individueller Prädiktoren von Erfolg auf der Resultateebene (vgl. Tabelle 54) ist zu beachten, dass es sich bei den eingeschlossenen Merkmalen hier wesentlich um gesundheitsbezogene und schichtbezogene Befindensaspekte handelt, die damit auf einer deutlich anderen Ebene angesiedelt sind als die vorherigen Punkte der Lern- und Verhaltensebene. Zudem ist zu beachten, dass insbesondere Alter, aber auch Geschlecht, für viele dieser gesundheitsbezogenen Aspekte differenzielle Befunde erwarten lassen. Vor diesem Hintergrund finden sich folgende individuellen Prädiktoren von Erfolg auf der Resultateebene: Wie auf der Lernebene sind auf dieser Ebene vor allem Frauen kurzfristig erfolgreich und tendenziell erneut Personen, die zu Beginn der Studie weniger Freude am Fahren mit Sondersignal angegeben haben. Bei den längerfristig erfolgreichen Teilnehmern auf der Resultateebene finden sich (im Gegensatz zu den vorherigen Ebenen) vor allem jüngere Personen und weniger erfahrene Personen mit geringerer Beschäftigungsdauer im Rettungsdienst, kürzerem Führerscheinbesitz, weniger Zeit als Fahrer im Rettungsdienst und tendenziell geringerer jährlicher Fahrleistung.

Im gewichteten Index zum Gesamterfolg werden alle Ebenen nach Relevanz für das Trainingsziel betrachtet. Die Unterschiede zwischen erfolgreichen und nichterfolgreichen Trainingsteilnehmern in diesem Index sind in Tabelle 55 dargestellt. Für den kurzfristigen Erfolg finden sich Unterschiede individueller Voraussetzungen ausschließlich in Merkmalen, die auch auf den anderen Ebenen relevant waren. So sind kurzfristig Gesamterfolgreiche älter, besitzen länger ihren Führerschein, fahren weniger RTW und haben weniger Spaß am Fahren mit Sondersignal. Für längerfristigen Gesamterfolg können keine individuellen Merkmalsunterschiede identifiziert werden – dies ist möglicherweise darin begründet, dass einige der in den Gesamtindex einfließenden Variablen, v.a. Alter und Erfahrung, sich längerfristig auch schon auf den einzelnen Ebenen unterschiedlich „verhalten“ haben und sich potenzielle Effekte im Gesamtindex damit gegenseitig aufheben könnten. Zudem wurde im Globalindex die Verhaltensebene besonders stark gewichtet, für die auch bei der Einzelbetrachtung der Ebene kaum individuelle Prädiktoren gefunden wurden.

Tabelle 54: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Resultateebene durch individuelle Voraussetzungen

Merkmal	Kurzfristig erfolgreich: Resultateebene, n ≤ 92	Kurzfristig nicht erfolgreich: Resultateebene, n ≤ 72	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Resultateebene, n ≤ 29	längerfristig nicht erfolgreich: Resultateebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Geschlecht			3.90	p=.060 <sup>+</sup>			0.18	p=.741 n.s.
Männer	66	61			22	17		
Frauen	26	11			7	4		
Alter	M = 33.6, SD = 9.27	M = 32.3, SD = 9.72	-1.20	p=.233 n.s.	M = 32.2, SD = 9.71	M = 37.8, SD = 10.0	-2.10	p=.036*
Qualifikation			2.03	p=.639 n.s.			1.45	p=.674
Rettungsdiensthelfer	3	1			0	1		
Rettungssanitäter	18	20			5	3		
Rettungsassistent	70	50			24	17		
Notfallsanitäter	1	1			0	0		
Dauer Beschäftigung im Rettungsdienst (Monate)	M = 137.5 SD = 109.02	M = 119.4 SD = 101.51	-1.14	p=.255 n.s.	M = 131.2 SD = 110.54	M = 171.3 SD = 109.57	-1.76	p=.079 <sup>+</sup>
Beschäftigungsart			0.15	p=.936 n.s.			0.15	p=1.00 n.s.
hauptberuflich	81	64			26	19		
nebenberuflich	5	3			2	1		
ehrenamtlich	6	5			1	1		
Bundesland			0.97	p=.409 n.s.			0.01	p=1.00 n.s.
Bayern	63	44			17	12		
Sachsen	29	28			12	9		
Anzahl Führerscheinklassen	M = 4.29, SD = 1.70	M = 4.36, SD = 1.58	-0.39	p=.702 n.s.	M = 4.34, SD = 1.72	M = 4.62, SD = 1.43	-0.68	p=.502 n.s.
Jahre Führerscheinbesitz	M = 15.0; SD = 8.74	M = 13.8; SD = 9.37	-1.14	p=.254 n.s.	M = 13.6; SD = 9.76	M = 18.7; SD = 8.68	-2.04	p=.041*
Erlaubnis zur Fahrgastbeförderung			2.86	p=.111 n.s.			0.01	p=1.00 n.s.
ja	14	18			8	6		
nein	76	50			21	15		
Jährliche Fahrleistung gesamt in km (alle Fahrzeugklassen)	M = 32473 SD = 15838.4	M = 30983 SD = 15047.1	-0.57	p=.568 n.s.	M = 31553 SD = 14092.3	M = 38005 SD = 14678.2	-1.64	p=.103 n.s.

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich: Resultatebene, n ≤ 92	Kurzfristig nicht erfolgreich: Resultatebene, n ≤ 72	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Resultatebene, n ≤ 29	längerfristig nicht erfolgreich: Resultatebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Punkte Fahreignungsregister			0.29	p=.637 n.s.			0.87	p=.388 n.s.
ja	48	41			17	15		
nein	43	31			12	6		
Jemals in Unfall mit Rettungswagen involviert?			0.13	p=.753 n.s.			0.67	p=.560 n.s.
ja	50	37			16	14		
nein	41	34			13	7		
Anzahl der Unfälle mit RTW als Fahrer	M = 0.55 SD = 0.50	M = 0.54 SD = 0.77	-1.04	p=.316 n.s.	M = 0.72 SD = 0.84	M = 0.67 SD = 0.48	-0.23	p=.847 n.s.
Stunden pro Monat als Fahrer auf RTW	M = 82.4 SD = 51.36	M = 81.1 SD = 51.03	-0.28	p=.784 n.s.	M = 73.5 SD = 50.51	M = 101.8 SD = 48.55	-1.98	p=.048*
„Freude am Fahren mit Sondersignal“ (Fahrer)	M = 3.54 SD = 0.99	M = 3.79 SD = 0.95	-1.60	p=.110 n.s.	M = 3.60 SD = 0.92	M = 3.30 SD = 1.26	-0.80	p=.431 n.s.
Beanspruchung durch Fahren mit SoSi (Fahrer)	M = 2.93 SD = 1.14	M = 2.84 SD = 1.12	-0.55	p=.584 n.s.	M = 2.91 SD = 1.04	M = 3.10 SD = 1.37	-0.38	p=.714 n.s.
Anzahl verkehrsrelevanter Fortbildungen			0.60	p=.896 n.s.			2.43	p=.488 n.s.
Keine	32	23			7	8		
1 - 2	46	40			19	12		
3 - 4	9	7			2	0		
5 und mehr	4	2			1	1		
Blaulichteinweisung innerhalb der letzten fünf Jahre?			0.06	p=.837 n.s.			0.15	p=1.00 n.s.
nein	16	13			4	4		
ja	71	52			23	17		
Bereits zu T1 guter Fahrer?			0.03	p=1.00 n.s.			0.87	p=.442 n.s.
nein	52	38			15	9		
ja	10	8			4	5		

<sup>1</sup> n.s. = nicht signifikant; Chi-Quadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

Tabelle 55: Prädiktion kurzfristigen und längerfristigen Gesamterfolgs durch individuelle Voraussetzungen

Merkmal	Kurzfristig erfolgreich: global, n ≤ 50	Kurzfristig nicht erfolgreich: global, n ≤ 82	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: global, n ≤ 13	längerfristig nicht erfolgreich: global, n ≤ 22	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Geschlecht			1.10	p=.380 n.s.			0.28	p=.689 n.s.
Männer	37	67			11	17		
Frauen	13	15			2	5		
Alter	M = 34.1, SD = 8.06	M = 32.1, SD = 10.0	-2.06	p=.039*	M = 33.4, SD = 7.91	M = 35.7, SD = 10.9	-0.38	p=.717 n.s.
Qualifikation			2.56	p=.514 n.s.			2.06	p=.471 n.s.
Rettungsdiensthelfer	2	2			0	1		
Rettungssanitäter	9	23			1	5		
Rettungsassistent	39	56			12	16		
Notfallsanitäter	0	1			0	0		
Dauer Beschäftigung im Rettungsdienst (Monate)	M = 134.8 SD = 97.06	M = 121.5 SD = 104.15	-1.13	p=.262 n.s.	M = 116.5 SD = 92.79	M = 163.9 SD = 108.17	-1.37	p=.178 n.s.
Beschäftigungsart			1.05	p=.702 n.s.			1.36	p=.771 n.s.
hauptberuflich	45	71			12	19		
nebenberuflich	3	4			0	2		
ehrenamtlich	2	7			1	1		
Bundesland			0.43	p=.568 n.s.			1.02	p=.481 n.s.
Bayern	32	57			6	14		
Sachsen	18	25			7	8		
Anzahl Führerscheinklassen	M = 4.20, SD = 1.58	M = 4.22, SD = 1.67	-0.01	p=.993 n.s.	M = 4.92, SD = 1.85	M = 4.27, SD = 1.32	-0.89	p=.384 n.s.
Jahre Führerscheinbesitz	M = 15.5; SD = 7.75	M = 13.6; SD = 9.55	-2.02	p=.043*	M = 15.2; SD = 8.05	M = 16.5; SD = 9.72	-0.27	p=.794 n.s.
Erlaubnis zur Fahrgastbeförderung			0.31	p=.663 n.s.			0.28	p=.698 n.s.
ja	12	17			4	5		
nein	35	63			9	17		
Jährliche Fahrleistung gesamt in km (alle Fahrzeugklassen)	M = 34273 SD = 16193.9	M = 30900 SD = 15268.8	-1.12	p=.264 n.s.	M = 36692 SD = 13960.6	M = 33931 SD = 14222.8	-0.32	p=.755 n.s.

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich: global, n ≤ 50	Kurzfristig nicht erfolgreich: global, n ≤ 82	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: global, n ≤ 13	längerfristig nicht erfolgreich: global, n ≤ 22	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Punkte Fahreignungsregister			0.08	p=.857 n.s.			0.36	p=.721 n.s.
ja	29	45			9	13		
nein	21	36			4	9		
Jemals in Unfall mit Rettungswagen involviert?			0.17	p=.720 n.s.			1.02	p=.481 n.s.
ja	25	43			6	14		
nein	25	37			7	8		
Anzahl der Unfälle mit RTW als Fahrer	M = 0.50 SD = 0.50	M = 0.54 SD = 0.50	-0.51	p=.596 n.s.	M = 0.46 SD = 0.52	M = 0.64 SD = 0.49	-1.04	p=.338 n.s.
Stunden pro Monat als Fahrer auf RTW	M = 75.2 SD = 47.95	M = 92.6 SD = 52.29	-1.73	p=.084 <sup>+</sup>	M = 85.3 SD = 42.11	M = 99.1 SD = 60.20	-0.71	p=.489 n.s.
„Freude am Fahren mit Sondersignal“ (Fahrer)	M = 3.33 SD = 0.90	M = 3.83 SD = 0.96	-2.93	p=.003**	M = 3.69 SD = 0.95	M = 3.29 SD = 1.15	-1.04	p=.309 n.s.
Beanspruchung durch Fahren mit SoSi (Fahrer)	M = 2.96 SD = 1.12	M = 2.93 SD = 1.11	-0.11	p=.914 n.s.	M = 3.00 SD = 1.22	M = 3.29 SD = 1.15	-0.73	p=.480 n.s.
Anzahl verkehrsrelevanter Fortbildungen			1.04	p=.793 n.s.			4.68	p=.197 n.s.
Keine	20	26			2	8		
1 - 2	24	43			9	14		
3 - 4	5	9			1	0		
5 und mehr	1	3			1	0		
Blaulichteinweisung innerhalb der letzten fünf Jahre?			0.10	p=.810 n.s.			0.01	p=1.00 n.s.
nein	9	13			2	4		
ja	37	62			10	18		
Bereits zu T1 guter Fahrer?			0.04	p=1.00 n.s.			1.06	p=.388 n.s.
nein	31	54			9	14		
ja	7	11			1	5		

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

## 6.4 Prädiktion von Erfolg durch organisationsbezogene Voraussetzungen

Analog zu der Darstellung der individuellen Voraussetzungen für Trainingserfolg werden in Tabelle 56 bis Tabelle 59 Analysen zu Unterschieden in den Arbeitsbedingungen zum ersten Messzeitpunkt erfolgreicher und nichterfolgreicher Teilnehmer präsentiert. Betrachtet werden dabei Anforderungen, Stressoren und Ressourcen in der Arbeit mit der Grundannahme, dass bessere Arbeitsbedingungen auch lernförderlich sind (vgl. Hacker, 2003, 2005). Allerdings ist denkbar, dass einzelne Bedingungen hier weniger relevant sind als Konstellationen von Bedingungen (vgl. Glaser, Seubert, Hornung & Herbig, 2015), was sich allerdings aufgrund der geringeren Stichprobenzahlen in diesen Erfolgsauswertungen nicht abbilden lässt.

Insgesamt finden sich tatsächlich nur wenige (durchgängige) und statistisch signifikante Unterschiede. Im Bereich des Erfolgs auf Lernebene (vgl. Tabelle 56) berichten sowohl kurz- wie langfristig erfolgreiche Trainingsteilnehmer höhere patientenbezogene Stressoren als nicht erfolgreiche Teilnehmer, allerdings sind die Ausprägungen dieses Stressoren allgemein auf einem eher geringen bis mittleren Niveau. Darüber hinaus geben nur für den längerfristigen Lernerfolg, erfolgreiche Personen einen höheren unspezifischen Zeitdruck bei der Arbeit und eine höhere Arbeitsplatzunsicherheit als nicht erfolgreiche an.

Patientenbezogenen Stressoren und Arbeitsplatzunsicherheit sowie spezifischer Zeitdruck unterscheiden sich auch zwischen auf der Verhaltensebene (vgl. Tabelle 57) kurzfristig erfolgreichen und nicht erfolgreichen Teilnehmern in der gleichen Richtung, d.h. Personen, die höhere Stressoren angeben, profitieren in ihrem Verhalten mehr von den Trainings. Ähnlich ist dieser Effekt auch für widersprüchliche Aufträge zu sehen. Für längerfristigen Erfolg auf der Verhaltensebene zeigt sich hingegen keiner dieser Unterschiede. Hier finden sich eher Unterschiede in den vorhandenen arbeitsbezogenen Ressourcen. So geben Teilnehmer, die ihr Verhalten längerfristig positiv verändert haben, deutlich mehr Rollenklarheit an, als die nicht erfolgreichen Teilnehmer. In etwas schwächerer Ausprägung geben die längerfristig Erfolgreichen auch eine höhere zeitliche Transparenz und bessere externe Kooperationen in ihrer Arbeit an.

Wie bei den längerfristig angelegten Merkmalen der Resultateebene zu erwarten, sind in diesem Bereich keine Unterschiede zwischen den kurzfristig erfolgreichen und nicht erfolgreichen Personen zu finden (vgl. Tabelle 58). Diejenigen, die längerfristige Verhaltensänderungen zeigen, geben einen geringeren spezifischen Zeitdruck an und – kontraintuitiv – eine geringere Führungsqualität und soziale Unterstützung durch ihre Vorgesetzten als die längerfristig nicht erfolgreichen Teilnehmer.

Ähnlich unerwartet fallen die Ergebnisse der Vergleiche für den Gesamterfolg aus (vgl. Tabelle 59). Als kurzfristig global erfolgreich kategorisierte Teilnehmer geben weniger Ressourcen im Sinne von Arbeitszeitautonomie, interne Kooperation und soziale Unterstützung durch Kollegen sowie höhere patientenbezogene Stressoren, höheren allgemeinen Zeitdruck und einen stärker strafenden Umgang mit Fehlern in ihren Organisationen an als nicht erfolgreiche Teilnehmer. Bei letzterem kehrt sich der Effekt für den längerfristigen Erfolg um – bei diesem Vergleich profitieren diejenigen Personen mit einer geringeren Ausprägung dieses Stressors insgesamt mehr von dem Training. Gleiches gilt für spezifischen Zeitdruck, auch hier ist der Stressor bei den längerfristig global erfolgreichen Teilnehmern geringer ausgeprägt als bei den nicht erfolgreichen.

Deskriptiv ist dieses Muster auch generell in den Vergleichen zum Gesamterfolg und zum Erfolg auf der Verhaltensebene zu finden – bei kurzfristig erfolgreichen finden sich eher ungünstige(re) Arbeitsbedingungskonstellationen (s.o.) mit höheren Stressoren und geringeren Ressourcen und eine Umkehrung dieser „Vorzeichen“ bei längerfristig erfolgreichen Personen mit geringeren Stressoren und höheren Ressourcen.

Tabelle 56: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Lernebene durch organisationsbezogene Voraussetzungen

Merkmal	Kurzfristig erfolgreich: Lernebene, n ≤ 91		Kurzfristig nicht erfolgreich: Lernebene, n ≤ 63		Z-Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Lernebene, n ≤ 24		längerfristig nicht erfolgreich: Lernebene, n ≤ 21		Z-Statistik	Unter- schied <sup>1</sup>
	M	SD	M	SD			M	SD	M	SD		
<b>Anforderungen</b>												
Geistige Anforderungen	4.09	0.583	4.19	0.557	-0.56	.576 n.s.	4.17	0.555	4.06	0.570	-0.78	.442 n.s.
Lernerfordernisse	4.10	0.622	4.11	0.644	-0.11	.912 n.s.	4.13	0.596	4.10	0.642	-0.10	.921 n.s.
Qualifizierungsmöglichkeiten	3.76	0.811	3.82	0.814	-0.42	.677 n.s.	3.61	0.664	3.57	0.920	-0.21	.841 n.s.
Emotionale Anforderungen	2.90	0.748	2.73	0.627	-1.62	.106 n.s.	3.00	0.918	2.90	0.659	-0.49	.630 n.s.
<b>Ressourcen</b>												
Zeitliche Transparenz/ Vorhersehbarkeit	3.07	0.600	2.95	0.708	-1.44	.151 n.s.	3.12	0.651	3.14	0.773	-0.40	.695 n.s.
Tätigkeitsspielräume	3.19	0.623	3.21	0.653	-0.52	.601 n.s.	3.05	0.632	3.15	0.707	-1.36	.179 n.s.
Arbeitszeitautonomie	2.17	0.679	2.42	0.923	-1.42	.157 n.s.	2.11	0.810	2.08	0.644	-0.08	.941 n.s.
Vorgesetztenfeedback	2.37	1.016	2.18	0.947	-1.00	.319 n.s.	2.33	0.843	2.12	1.011	-0.84	.411 n.s.
Partizipationschancen	2.27	0.712	2.31	0.799	-0.54	.592 n.s.	2.11	0.611	2.16	0.860	-0.20	.847 n.s.
Räumliche Ressourcen	3.04	0.1049	3.23	1.220	-1.09	.278 n.s.	2.58	0.945	3.01	1.237	-1.09	.279 n.s.
Materielle Ressourcen	3.71	0.763	3.69	0.803	-0.13	.900 n.s.	3.61	0.931	3.55	0.881	-0.21	.843 n.s.
Rollenklarheit	4.15	0.548	4.05	0.704	-0.47	.639 n.s.	4.16	0.459	3.89	0.761	-1.05	.302 n.s.
Führungsqualität	2.99	0.964	2.98	0.910	-0.18	.856 n.s.	2.90	0.903	2.82	0.992	-0.35	.730 n.s.
Interne Kooperation	3.84	0.704	3.93	0.671	-0.75	.458 n.s.	3.53	0.661	3.79	0.815	-0.91	.370 n.s.
Externe Kooperation	3.70	0.442	3.75	0.484	-0.41	.683 n.s.	3.59	0.350	3.62	0.419	-0.22	.833 n.s.
Soziale Unterstützung Vorgesetzte	3.16	1.062	3.31	1.022	-0.73	.466 n.s.	3.01	1.044	2.90	0.995	-0.73	.474 n.s.
Soziale Unterstützung Kollegen	3.70	0.850	3.79	0.817	-0.80	.428 n.s.	3.40	0.821	3.76	0.995	-1.54	.126 n.s.
<b>Stressoren</b>												
Organisationale Stressoren	3.09	0.711	3.09	0.693	-0.10	.920 n.s.	3.34	0.677	3.12	0.741	-1.08	.284 n.s.
Patientenbezogene Stressoren	2.97	0.631	2.73	0.756	-2.29	.022*	3.23	0.662	2.84	0.887	-1.73	.084 <sup>†</sup>

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich: Lernebene, n ≤ 91		Kurzfristig nicht erfolgreich: Lernebene, n ≤ 63		Z-Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Lernebene, n ≤ 24		längerfristig nicht erfolgreich: Lernebene, n ≤ 21		Z-Statistik	Unter- schied <sup>1</sup>
	M	SD	M	SD			M	SD	M	SD		
Zeitdruck bei unspezifischen Festlegungen	2.82	0.780	2.64	0.748	-0.97	.333 n.s.	3.00	1.000	2.50	0.548	-1.76	.079 <sup>+</sup>
Zeitdruck bei spezifischen Festlegungen	2.32	0.565	2.24	0.605	-0.68	.498 n.s.	2.42	0.660	2.19	0.705	-1.06	.294 n.s.
Widersprüchliche Aufträge	2.29	0.656	2.33	0.687	-0.66	.512 n.s.	2.38	0.684	2.31	0.753	-0.22	.834 n.s.
Informativische Erschwernisse	3.14	0.643	3.13	0.673	-0.14	.893 n.s.	3.39	0.725	3.16	0.728	-1.21	.230 n.s.
Arbeitsunterbrechungen	2.51	0.541	2.46	0.588	-0.22	.824 n.s.	2.56	0.622	2.33	0.477	-1.58	.117 n.s.
Qualitätseinbußen	2.22	0.662	2.12	0.791	-0.97	.334 n.s.	2.31	0.715	2.24	0.761	-0.32	.754 n.s.
Ungünstige Arbeitsumgebung (Einsatzort)	3.58	0.577	3.44	0.597	-1.55	.123 n.s.	3.65	0.690	3.53	0.588	-0.54	.599 n.s.
Ungünstige Arbeitsumgebung (RTW)	2.56	0.548	2.57	0.681	-0.20	.841 n.s.	2.62	0.698	2.61	0.789	-0.50	.623 n.s.
Strafender Umgang mit Fehlern	2.49	0.791	2.41	0.824	-0.61	.547 n.s.	2.62	0.724	2.60	1.063	-0.48	.636 n.s.
Arbeitsplatzunsicherheit	1.74	0.643	1.72	0.621	-0.07	.945 n.s.	2.14	0.886	1.59	0.427	-1.93	.054 <sup>+</sup>

<sup>1</sup> n.s. = nicht signifikant; Chi-Quadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

Tabelle 57: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Verhaltensebene durch organisationsbezogene Voraussetzungen

Merkmal	Kurzfristig erfolgreich Verhaltensebene, n ≤ 64		Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 54		Z-Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤ 16		längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14		Z-Statistik	Unter- schied <sup>1</sup>
	M	SD	M	SD			M	SD	M	SD		
Anforderungen												
Geistige Anforderungen	4.12	0.566	4.15	.538	-0.12	.904 n.s.	4.22	.507	3.96	.571	-0.97	.342 n.s.
Lernerfordernisse	4.05	0.608	4.15	.647	-1.12	.266 n.s.	4.23	.767	4.17	.408	-0.63	.546 n.s.
Qualifizierungsmöglichkeiten	3.77	0.792	3.83	.717	-0.40	.690 n.s.	3.79	.749	3.57	.789	-0.42	.687 n.s.
Emotionale Anforderungen	2.88	0.726	2.75	.615	-0.67	.506 n.s.	2.83	.925	2.70	.804	-0.23	.829 n.s.



Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich Verhaltensebene, n ≤ 64		Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 54		Z-Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤ 16		längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14		Z-Statistik	Unter- schied <sup>1</sup>
Ressourcen												
Zeitliche Transparenz/ Vorhersehbarkeit	3.06	0.617	3.01	0.671	-0.23	.823 n.s.	3.26	0.722	2.82	.675	-1.76	.081 <sup>+</sup>
Tätigkeitsspielräume	3.22	0.561	3.20	0.715	-0.36	.722 n.s.	3.21	0.578	3.09	.887	-0.02	.992 n.s.
Arbeitszeitautonomie	2.26	0.783	2.39	0.856	-0.71	.478 n.s.	2.02	0.541	2.20	.736	-0.49	.640 n.s.
Vorgesetztenfeedback	2.26	0.955	2.34	1.050	-0.41	.683 n.s.	2.38	1.190	2.11	.924	-0.45	.665 n.s.
Partizipationschancen	2.29	0.660	2.36	0.851	-0.18	.863 n.s.	2.23	0.909	2.21	.747	-0.04	.975 n.s.
Räumliche Ressourcen	3.04	1.030	3.28	1.265	-1.22	.223 n.s.	2.74	1.092	3.13	1.194	-1.06	.297 n.s.
Materielle Ressourcen	3.60	0.808	3.84	0.720	-1.63	.103 n.s.	3.85	0.655	3.36	1.041	-1.07	.295 n.s.
Rollenklarheit	4.10	0.636	4.11	0.577	-0.09	.929 n.s.	4.45	0.425	3.77	.550	-3.09	.001***
Führungsqualität	2.95	0.928	3.00	0.984	-0.26	.798 n.s.	2.86	0.959	2.99	.882	-0.31	.771 n.s.
Interne Kooperation	3.73	0.734	4.00	0.638	-2.00	.045*	3.61	0.626	3.79	.825	-0.81	.430 n.s.
Externe Kooperation	3.72	0.406	3.82	0.478	-1.28	.201 n.s.	3.86	0.338	3.57	.447	-1.67	.097 <sup>+</sup>
Soziale Unterstützung Vorgesetzte	3.13	1.004	3.17	1.078	-0.24	.813 n.s.	3.02	1.127	2.96	.877	-0.02	1.00 n.s.
Soziale Unterstützung Kollegen	3.68	0.847	3.69	0.814	-0.05	.959 n.s.	3.47	1.103	3.71	.802	-0.44	.672 n.s.
Stressoren												
Organisationale Stressoren	3.10	0.735	3.08	0.689	-0.15	.885 n.s.	3.12	0.705	3.18	0.801	-0.08	.943 n.s.
Patientenbezogene Stressoren	3.05	0.743	2.68	0.565	-3.07	.002**	2.86	0.907	3.02	0.809	-0.17	.878 n.s.
Zeitdruck bei unspezifischen Festlegungen	2.89	0.838	2.47	0.610	-2.45	.014*	2.63	0.922	2.68	0.823	-0.09	.942 n.s.
Zeitdruck bei spezifischen Festlegungen	2.33	0.614	2.19	0.515	-1.16	.247 n.s.	2.20	0.702	2.45	0.575	-1.38	.172 n.s.
Widersprüchliche Aufträge	2.43	0.647	2.22	0.681	-1.65	.099 <sup>+</sup>	2.39	0.844	2.29	0.591	-0.63	.543 n.s.
Informativische Erschwernisse	3.14	0.692	3.13	.628	-0.17	.870 n.s.	3.31	0.821	3.19	0.535	-0.61	.557 n.s.
Arbeitsunterbrechungen	2.54	0.590	2.38	0.485	-1.01	.313 n.s.	2.31	0.661	2.61	0.507	-1.56	.123 n.s.

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich Verhaltensebene, n ≤ 64		Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 54		Z-Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤ 16		längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14		Z-Statistik	Unter- schied <sup>1</sup>
	M	SD	M	SD			M	SD	M	SD		
Qualitätseinbußen	2.26	0.804	2.11	0.647	-0.58	.562 n.s.	2.17	.798	2.40	0.764	-0.65	.527 n.s.
Ungünstige Arbeitsum- gebung (Einsatzort)	3.49	0.573	3.53	0.597	-0.52	.604 n.s.	3.55	.546	3.67	0.701	-0.21	.845 n.s.
Ungünstige Arbeitsumgebung (RTW)	2.54	0.568	2.51	0.607	-0.39	.700 n.s.	2.32	0.622	2.81	0.825	-1.63	.107 n.s.
Strafender Umgang mit Fehlern	2.50	0.763	2.45	0.896	-0.46	.648 n.s.	2.60	1.063	2.71	0.815	-0.71	.490 n.s.
Arbeitsplatzunsicherheit	1.92	0.661	1.62	0.574	-2.51	.012*	1.95	0.642	1.79	0.592	-0.42	.691 n.s.

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveaus: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

Tabelle 58: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Resultateebene durch organisationsbezogene Voraussetzungen

Merkmal	Kurzfristig erfolg- reich Resultate- ebene, n ≤ 91		Kurzfristig nicht erfolgreich: Resul- tateebene, n ≤ 71		Z-Statistik	Unter- schied <sup>1</sup>	längerfristig erfolgreich: Resul- tateebene, n ≤ 29		längerfristig nicht erfolgreich: Resul- tateebene, n ≤ 21		Z-Statistik	Unter- schied <sup>1</sup>
	M	SD	M	SD			M	SD	M	SD		
<b>Anforderungen</b>												
Geistige Anforderungen	4.14	0.552	4.13	.581	-0.07	.947 n.s.	4.03	0.516	4.27	0.612	-1.36	.177 n.s.
Lernerfordernisse	4.14	0.577	4.04	.666	-0.62	.537 n.s.	4.08	0.665	4.25	0.536	-0.77	.451 n.s.
Qualifizierungsmöglichkeiten	3.76	0.800	3.79	.811	-0.32	.751 n.s.	3.57	0.796	3.76	0.824	-0.56	.583 n.s.
Emotionale Anforderungen	2.80	0.741	2.88	.632	-0.60	.547 n.s.	2.81	0.712	3.01	0.927	-0.36	.728 n.s.
<b>Ressourcen</b>												
Zeitliche Transparenz/ Vorhersehbarkeit	2.99	0.690	3.04	0.619	-0.52	.603 n.s.	3.20	0.635	2.95	0.725	-1.27	.209 n.s.
Tätigkeitsspielräume	3.23	0.608	3.18	0.644	-0.40	.690 n.s.	3.16	0.517	3.10	0.877	-0.27	.795 n.s.
Arbeitszeitautonomie	2.25	0.714	2.29	0.865	-0.26	.797 n.s.	2.08	0.706	2.12	0.746	-0.01	.996 n.s.
Vorgesetztenfeedback	2.36	0.972	2.17	0.982	-1.25	.213 n.s.	2.14	0.925	2.45	1.048	-1.00	.321 n.s.
Partizipationschancen	2.32	.716	2.24	0.750	-0.87	.388 n.s.	2.22	0.714	2.02	0.771	-0.87	.389 n.s.
Räumliche Ressourcen	3.18	1.154	3.07	1.075	-0.59	.558 n.s.	2.74	0.983	2.76	1.174	-0.32	.758 n.s.

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich Resultateebene, n ≤ 91		Kurzfristig nicht erfolgreich: Resultateebene, n ≤ 71		Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Resultateebene, n ≤ 29		längerfristig nicht erfolgreich: Resultateebene, n ≤ 21		Z-Statistik	Unterschied <sup>1</sup>
Materielle Ressourcen	3.67	0.812	3.73	0.738	-0.31	.761 n.s.	3.53	1.004	3.63	0.685	-0.03	.981 n.s.
Rollenklarheit	4.06	0.642	4.14	0.565	-0.75	.457 n.s.	4.09	0.524	4.06	0.765	-0.22	.835 n.s.
Führungsqualität	2.98	0.884	2.97	1.002	-0.02	.988 n.s.	2.67	0.816	3.18	0.977	-2.17	.029*
Interne Kooperation	3.82	0.697	3.91	0.668	-0.83	.406 n.s.	3.59	0.695	3.82	0.848	-0.80	.429 n.s.
Externe Kooperation	3.75	.446	3.67	0.463	-1.29	.197 n.s.	3.64	0.369	3.68	0.451	-0.27	.794 n.s.
Soziale Unterstützung Vorgesetzte	3.26	1.018	3.20	1.061	-0.30	.763 n.s.	2.75	0.877	3.37	1.182	-1.94	.052 <sup>†</sup>
Soziale Unterstützung Kollegen	3.68	0.840	3.80	0.795	-0.81	.417 n.s.	3.57	0.832	3.60	1.080	-0.14	.894 n.s.
<b>Stressoren</b>												
Organisationale Stressoren	3.07	0.686	3.13	0.732	-0.51	.609 n.s.	3.20	0.786	3.15	0.614	-0.49	.629 n.s.
Patientenbezogene Stressoren	2.85	0.719	2.91	0.666	-0.40	.688 n.s.	2.85	0.685	3.23	0.873	-1.46	.148 n.s.
Zeitdruck bei unspezifischen Festlegungen	2.73	0.768	2.75	0.783	-0.01	.992 n.s.	2.66	0.721	2.98	1.006	-1.24	.217 n.s.
Zeitdruck bei spezifischen Festlegungen	2.26	0.551	2.32	.604	-0.48	.634 n.s.	2.11	0.537	2.55	0.733	-2.23	.025*
Widersprüchliche Aufträge	2.30	.670	2.36	0.662	-0.14	.890 n.s.	2.43	0.754	2.20	0.713	-0.90	.375 n.s.
Informatorische Erschwernisse	3.11	0.645	3.17	0.656	-0.63	.529 n.s.	3.30	0.730	3.20	0.750	-0.38	.707 n.s.
Arbeitsunterbrechungen	2.49	0.565	2.48	0.531	-0.24	.810 n.s.	2.50	0.448	2.37	0.655	-0.99	.330 n.s.
Qualitätseinbußen	2.20	0.713	2.17	.728	-0.17	.869 n.s.	2.36	0.771	2.11	0.710	-1.12	.266 n.s.
Ungünstige Arbeitsumgebung (Einsatzort)	3.56	0.621	3.48	0.562	-0.71	.479 n.s.	3.68	0.647	3.43	0.573	-1.39	.168 n.s.
Ungünstige Arbeitsumgebung (RTW)	2.60	0.594	2.54	0.599	-1.03	.306 n.s.	2.60	0.679	2.56	0.778	-0.48	.636 n.s.
Strafender Umgang mit Fehlern	2.54	0.786	2.39	0.809	-1.57	.118 n.s.	2.60	0.870	2.55	0.932	-0.21	.843 n.s.
Arbeitsplatzunsicherheit	1.76	0.587	1.69	0.675	-1.15	.252 n.s.	1.69	.610	2.01	0.856	-1.15	.255 n.s.

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: <sup>†</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

Tabelle 59: Prädiktion kurzfristigen und längerfristigen Gesamterfolgs durch organisationsbezogene Voraussetzungen

Merkmal	Kurzfristig erfolgreich global, n ≤ 50		Kurzfristig nicht erfolgreich: global, n ≤ 80		Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: global, n ≤ 13		längerfristig nicht erfolgreich: global, n ≤ 22		Z-Statistik	Unterschied <sup>1</sup>
	M	SD	M	SD			M	SD	M	SD		
<b>Anforderungen</b>												
Geistige Anforderungen	3.99	0.596	4.20	0.519	-1.60	.111 n.s.	4.08	0.504	4.09	0.595	-0.05	.966 n.s.
Lernerfordernisse	4.07	0.561	4.06	0.677	-0.29	.775 n.s.	4.13	0.845	4.09	0.556	-0.29	.778 n.s.
Qualifizierungsmöglichkeiten	3.71	0.945	3.81	0.721	-0.35	.730 n.s.	3.64	1.023	3.58	0.764	-0.54	.603 n.s.
Emotionale Anforderungen	2.89	0.787	2.80	0.613	-0.59	.561 n.s.	2.73	0.665	2.94	0.970	-0.65	.525 n.s.
<b>Ressourcen</b>												
Zeitliche Transparenz/ Vorhersehbarkeit	3.03	0.585	2.98	0.691	-0.27	.791 n.s.	3.01	0.638	3.06	0.748	-0.19	.859 n.s.
Tätigkeitsspielräume	3.28	0.563	3.14	0.683	-0.91	.365 n.s.	2.96	0.624	3.14	0.813	-1.13	.265 n.s.
Arbeitszeitautonomie	2.10	.650	2.45	0.919	-2.21	.027*	2.27	0.949	2.06	0.722	-0.78	.448 n.s.
Vorgesetztenfeedback	2.44	1.003	2.18	1.023	-1.42	.157 n.s.	2.27	1.285	2.09	0.921	-0.14	.896 n.s.
Partizipationschancen	2.29	0.741	2.32	0.768	-0.04	.970 n.s.	2.23	0.738	2.12	0.888	-0.66	.521 n.s.
Räumliche Ressourcen	3.07	1.118	3.18	1.131	-0.53	.595 n.s.	2.65	0.801	2.97	1.240	-0.98	.338 n.s.
Materielle Ressourcen	3.55	0.937	3.79	0.697	-1.40	.163 n.s.	3.86	1.090	3.41	0.893	-1.58	.117 n.s.
Rollenklarheit	4.11	0.598	4.09	0.623	0.00	.997 n.s.	4.08	0.597	4.07	0.637	-0.16	.879 n.s.
Führungsqualität	2.90	1.029	2.98	0.910	-0.55	.585 n.s.	2.71	0.871	2.79	1.037	-0.45	.662 n.s.
Interne Kooperation	3.70	0.808	3.94	0.614	-1.68	.093 <sup>+</sup>	3.48	0.767	3.66	0.762	-0.91	.375 n.s.
Externe Kooperation	3.71	0.437	3.78	0.435	-0.68	.499 n.s.	3.71	0.418	3.66	0.416	-0.33	.752 n.s.
Soziale Unterstützung Vorgesetzte	3.05	1.144	3.22	1.014	-0.93	.355 n.s.	2.83	1.096	2.86	1.039	-0.05	.959 n.s.
Soziale Unterstützung Kollegen	3.51	0.872	3.78	0.799	-1.83	.067 <sup>+</sup>	3.38	0.893	3.57	1.004	-0.61	.556 n.s.
<b>Stressoren</b>												
Organisationale Stressoren	3.10	0.737	3.09	0.702	-0.12	.910 n.s.	3.25	0.842	3.22	0.692	-0.34	.742 n.s.
Patientenbezogene Stressoren	3.06	0.627	2.75	0.678	-2.57	.010*	2.82	0.784	3.04	0.848	-0.22	.833 n.s.

Abschlussbericht „Achtung Blaulicht“

Merkmal	Kurzfristig erfolgreich global, n ≤ 50		Kurzfristig nicht erfolgreich: global, n ≤ 80		Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: global, n ≤ 13		längerfristig nicht erfolgreich: global, n ≤ 22		Z-Statistik	Unterschied <sup>1</sup>
Zeitdruck bei unspezifischen Festlegungen	2.85	0.749	2.63	0.787	-1.65	.099 <sup>+</sup>	2.46	0.721	2.77	1.009	-0.90	.378 n.s.
Zeitdruck bei spezifischen Festlegungen	2.29	0.545	2.25	0.604	-0.37	.713 n.s.	1.92	0.542	2.53	0.719	-2.71	.006**
Widersprüchliche Aufträge	2.43	0.691	2.27	0.691	-1.28	.203 n.s.	2.22	0.968	2.41	0.697	-0.91	.373 n.s.
Informatorische Erschwernisse	3.10	0.676	3.15	0.663	-0.44	.662 n.s.	3.16	0.828	3.35	0.691	-0.43	.679 n.s.
Arbeitsunterbrechungen	2.57	0.572	2.44	0.555	-1.08	.282 n.s.	2.44	0.480	2.43	0.651	-0.02	.994 n.s.
Qualitätseinbußen	2.29	0.823	2.14	0.683	-0.74	.463 n.s.	2.21	0.800	2.26	0.748	-0.07	.949 n.s.
Ungünstige Arbeitsumgebung (Einsatzort)	3.52	0.572	3.48	0.573	-0.47	.643 n.s.	3.56	0.548	3.58	0.649	-0.15	.886 n.s.
Ungünstige Arbeitsumgebung (RTW)	2.50	0.518	2.53	0.636	-0.02	.982 n.s.	2.32	0.586	2.67	0.813	-1.47	.145 n.s.
Strafender Umgang mit Fehlern	2.59	0.839	2.38	0.797	-1.67	.096 <sup>+</sup>	2.31	0.834	2.86	0.958	-1.70	.092 <sup>+</sup>
Arbeitsplatzunsicherheit	1.81	0.650	1.70	0.616	-0.81	.419 n.s.	1.65	0.391	2.05	0.802	-1.31	.197 n.s.

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveaus: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001

## 6.5 Prädiktion von Erfolg durch gruppenbezogene Voraussetzungen

Um ggf. Empfehlungen zum Training geben zu können, werden im Folgenden einige Voraussetzungen zu den Trainingsgruppen genauer betrachtet. Da die Voraussetzungen für Mehrebenenanalysen (zum Teil zu wenig Personen auf Ebene 1) nicht gegeben sind, werden für die Analysen Variablen auf Ebene der Trainingsgruppen gebildet und den einzelnen Probanden zugeordnet. Um dabei eine zu starke „Homogenisierung“ der Daten zu vermeiden, werden größtenteils Kategorisierungen verwendet. Im Einzelnen werden zunächst Aspekte untersucht, die nicht im Vorfeld eines Trainings änderbar sind: Die individuelle Bewertung der Gruppenatmosphäre und des gesamten Trainings (Schulnote) sowie als Unterpunkte, die hier zusätzlichen Einfluss auf den Trainingserfolg haben könnten, die jeweils durchschnittlichen Bewertungen der Atmosphäre und des Trainings durch die Gruppe und deren „Einigkeit“ in der Bewertung, operationalisiert über die Standardabweichung innerhalb der Gruppe. Zu den Variablen, die bereits im Vorfeld beachtet werden könnten, gehören Anzahl der Trainingsteilnehmer, Zusammensetzung der Trainingsgruppen im Hinblick auf Alter, Geschlecht, Rettungsdienst- und Fahrerfahrung sowie als Sonderpunkt, da sich dieser Aspekt auch in den individuellen Merkmalen zum Trainingserfolg als interessant erwiesen hat, die Freude am Fahren mit Sondersignal. Basierend auf Kategorien dieser Aspekte wurden die Trainingsgruppen jeweils als homogen oder heterogen, z.T. in unterschiedlichen Kombinationen klassifiziert und diese Klassifikation den jeweiligen Trainingsteilnehmern als „Merkmal“ zugeschrieben, um die Analysen wie in den vorherigen Abschnitten durchführen zu können. Im Einzelnen wurden die Alterskategorien bis 30 Jahren (jung), 30-50 (mittelalt) und ab 51 Jahren (alt) verwendet. Für den Erfahrungsbereich (sowohl beruflich wie fahrerisch) wurde die in der Expertiseforschung gängige Einteilung in Novizen, Intermediates und Experten verwendet (vgl. Herbig & Glöckner, 2009), und vereinfacht so definiert: Novizen sind alle Personen mit weniger als 5 Jahren Erfahrung im jeweiligen Bereich, Intermediates Personen zwischen 5 und 10 Jahren Erfahrungen, Experten Personen mit mehr als 10 Jahren Erfahrung. Als letzter Bereich wurde das einfache fünfstufige Item zum Spaß am Fahren mit Sondersignal in die Kategorien wenig Spaß (Werte 1 und 2), durchschnittlich Spaß (Wert 3) und viel Spaß (Werte 4 und 5) aufgeteilt.

In den dargestellten Ergebnissen (Tabelle 60 bis Tabelle 63) ist zu beachten, dass nicht alle Trainingsgruppen in den Analysen zum längerfristigen Erfolg vertreten sind – dies ist nicht so zu interpretieren, dass die entsprechenden Gruppenvariablen zu einem systematischen Ablehnen der letzten Messung geführt haben, sondern liegt an dem organisatorischen Ablauf der Studie. Teilnehmer von Trainings, die relativ spät in der Studie bzw. am Erhebungsstandort durchgeführt wurden, konnten aufgrund zu kurzer Abstände in den Messungen nicht in die letzte Messung aufgenommen werden (vgl. Abschnitt 3.4.2).

Auf der Lernebene sind kurzfristig Teilnehmer von Trainingsgruppen (TG) erfolgreich, in denen die Trainingsatmosphäre sowohl individuell als auch im Gruppenmittel besser bewertet wurde und diese Bewertung auch homogener war. Tendenziell fand sich dies auch für die Gesamtbewertung des Trainings. Darüber hinaus zeigte sich ein deutlicher Unterschied hinsichtlich der Rettungsdiensterfahrung – kurzfristig auf der Lernebene erfolgreiche Teilnahme stammten mehrheitlich aus Gruppen, die vollständig heterogen zusammengesetzt waren oder Kombinationen von jeweils zwei direkt aufeinanderfolgenden Expertisegrade hatten, während bei den nichterfolgreichen Teilnehmern eine Reihe von Personen aus Gruppen mit Novizen-Experten-Kombinationen stammten. Auch für den Spaß am Sondersignalfahren zeigten sich Unterschiede – wesentlich für erfolgreiche Teilnahme auf der Lernebene scheint erneut Heterogenität in der Trainingszusammensetzung und die Teilnahme von weniger gern fahrenden Personen zu sein. Für den längerfristigen Erfolg auf der Lernebene finden sich keinerlei systematischen Unterschiede in den Gruppenvariablen.

Tabelle 60: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Lernebene durch gruppenbezogene Voraussetzungen

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: Lernebene, n ≤ 91	Kurzfristig nicht erfolgreich: Lernebene, n ≤ 63	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Lernebene, n ≤ 24	längerfristig nicht erfolgreich: Lernebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Individuelle Bewertung der Atmosphäre in TG	M=4.96; SD = 0.206	M=4.79; SD = 0.481	-2.71	.008**	M=4.79; SD = 0.509	M=4.95; SD = 0.218	-1.27	.315 n.s.
Gruppenbewertung der Atmosphäre (M/Gruppe)	M=4.92; SD = 0.170	M=4.86; SD = 0.236	-2.00	.046*	M=4.88; SD = 0.286	M=4.89; SD = 0.229	-0.39	.728 n.s.
Homogenität dieser Bewertung (SD/Gruppe)	M=0.14; SD = 0.256	M=0.21; SD = 0.262	-1.91	.057 <sup>+</sup>	M=0.13; SD = 0.237	M=0.16; SD = 0.237	-0.39	.728 n.s.
Individuelle Note für Training	M=1.56; SD = 0.625	M=1.87; SD = 0.963	-1.91	.057 <sup>+</sup>	M=1.85; SD = 1.193	M=1.81; SD = 0.928	-0.44	.679 n.s.
Gruppennote (M/Gruppe)	M=1.64; SD = 0.496	M=1.83; SD = 0.833	-1.41	.160 n.s.	M=1.91; SD = 1.056	M=1.76; SD = 0.848	-0.37	.721 n.s.
Homogenität in der Note (SD/Gruppe)	M=0.55; SD = 0.202	M=0.55; SD = 0.272	-0.80	.426 n.s.	M=0.55; SD = 0.166	M=0.58; SD = 0.237	-0.56	.581 n.s.
Anzahl Trainingsteilnehmer			2.80	.225 n.s.			0.43	.552 n.s.
sehr wenige (1-3)	2	4			0	0		
wenige (4-5)	53	30			16	12		
wie geplant (6)	36	29			8	9		
Zusammensetzung: Geschlecht			1.02	.799 n.s.			1.95	.356 n.s.
ausschließlich männlich	25	20			10	6		
überwiegend männlich	54	36			13	15		
ausgeglichen	3	3			1	0		
überwiegend weiblich	9	4			0	0		
Zusammensetzung: Alter			0.45	.917 n.s.			1.81	.267 n.s.
Homogen jung	3	3			0	0		
Homogen mittelalt	3	3			0	0		
Heterogen jede Altersgruppe	23	15			3	6		
Heterogen jung + mittelalt	62	42			21	15		
Zusammensetzung: Rettungsdiensterfahrung			11.31	.026*			0.37	.689 n.s.
homogen Novizen	1	1			0	0		
homogen Intermediates	1	1			0	0		

Abschlussbericht „Achtung Blaulicht“

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: Lernebene, n ≤ 91	Kurzfristig nicht erfolgreich: Lernebene, n ≤ 63	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Lernebene, n ≤ 24	längerfristig nicht erfolgreich: Lernebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
heterogen alle Erfahrungsstufen	70	36			21	17		
heterogen Novizen + Intermediates	2	2			0	0		
heterogen Intermediates + Experten	6	2			0	0		
heterogen Novizen + Experten	11	21			3	4		
Zusammensetzung: Fahrerfahrung			3.08	.708 n.s.			2.93	.447 n.s.
homogen Intermediates	3	3			0	0		
Homogen Experten	16	13			3	6		
heterogen alle Erfahrungsstufen	32	17			9	7		
heterogen Novizen + Intermediates	1	2			0	0		
heterogen Intermediates + Experten	32	20			8	7		
heterogen Novizen + Experten	7	8			4	1		
Zusammensetzung: Spaß beim Sondersignalfahren			10.60	.058 <sup>+</sup>			3.48	.700 n.s.
Homogen: durchschnittlich Spaß	7	4			1	1		
Homogen: viel Spaß	25	22			5	5		
Vollständig heterogen	28	17			6	9		
Heterogen: wenig und durchschnittlich in vergleichbaren Anteilen	13	1			5	1		
Heterogen: durchschnittlich und viel in vergleichbaren Anteilen	10	14			6	4		
Heterogen: wenig und viel in vergleichbaren Anteilen	8	5			1	1		

<sup>1</sup> n.s. = nicht signifikant; Chi-Quadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveaus: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001; <sup>2</sup> für Kreuztabellen werden ausschließlich vorhandene Ausprägungen angegeben; Erläuterung der Kategorien im Text



Tabelle 61: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Verhaltensebene durch gruppenbezogene Voraussetzungen

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: Verhaltensebene, n ≤ 65	Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 55	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤ 16	längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Individuelle Bewertung der Atmosphäre in TG	M=4.85; SD = 0.441	M=4.91; SD = 0.290	-0.61	.501 n.s.	M=4.94; SD = 0.250	M=4.64; SD = 0.633	-1.64	.135 n.s.
Gruppenbewertung der Atmosphäre (M/Gruppe)	M=4.86; SD = 0.300	M=4.93; SD = 0.128	-0.40	.690 n.s.	M=4.92; SD = 0.254	M=4.79; SD = 0.353	-1.71	.088 <sup>+</sup>
Homogenität dieser Bewertung (SD/Gruppe)	M=0.16; SD = 0.266	M=0.14; SD = 0.235	-0.37	.718 n.s.	M=0.07; SD = 0.195	M=0.22; SD = 0.273	-1.71	.088 <sup>+</sup>
Individuelle Note für Training	M=1.89; SD = 1.138	M=1.54; SD = 0.686	-1.69	.091 <sup>+</sup>	M=1.59; SD = 1.020	M=2.07; SD = 1.328	-1.32	.202 n.s.
Gruppennote (M/Gruppe)	M=1.90; SD =1.055	M=1.65; SD = 0.412	-0.03	.974 n.s.	M=1.78; SD = 0.983	M=2.02; SD = 1.333	-0.36	.734 n.s.
Homogenität in der Note (SD/Gruppe)	M=0.54; SD = 0.173	M=0.55; SD = 0.239	-0.08	.936 n.s.	M=0.51; SD = 0.119	M=0.47; SD = 0.051	-0.93	.364 n.s.
Anzahl Trainingsteilnehmer			1.18	.597 n.s.			0.62	.484 n.s.
sehr wenige (1-3)	2	2			0	0		
wenige (4-5)	36	25			8	5		
wie geplant (6)	27	28			8	9		
Zusammensetzung: Geschlecht			1.66	.657 n.s.			0.98	1.00 n.s.
ausschließlich männlich	20	17			5	4		
überwiegend männlich	40	30			10	10		
ausgeglichen	2	4			1	0		
überwiegend weiblich	3	4			0	0		
Zusammensetzung: Alter			1.95	.690 n.s.			2.92	.228 n.s.
Homogen jung	2	3			0	0		
Homogen mittelalt	1	3			0	0		
Heterogen jede Altersgruppe	15	11			3	0		
Heterogen jung + mittelalt	47	38			13	14		
Zusammensetzung: Rettungsdienstenerfahrung			1.52	.919 n.s.			0.24	1.00 n.s.
homogen Novizen	1	1			0	0		
homogen Intermediates	1	1			0	0		

Abschlussbericht „Achtung Blaulicht“

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: Verhaltensebene, n ≤ 65	Kurzfristig nicht erfolgreich: Verhaltensebene, n ≤ 55	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Verhaltensebene, n ≤ 16	längerfristig nicht erfolgreich: Verhaltensebene, n ≤ 14	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
heterogen alle Erfahrungsstufen	48	40			14	13		
heterogen Novizen + Intermediates	1	2			0	0		
heterogen Intermediates + Experten	5	2			0	0		
heterogen Novizen + Experten	9	9			2	1		
Zusammensetzung: Fahrerfahrung			1.43	.930 n.s.			3.31	.434 n.s.
homogen Intermediates	2	3			0	0		
Homogen Experten	12	9			5	1		
heterogen alle Erfahrungsstufen	22	17			6	8		
heterogen Novizen + Intermediates	1	1			0	0		
heterogen Intermediates + Experten	24	19			4	3		
heterogen Novizen + Experten	4	6			1	2		
Zusammensetzung: Spaß beim Sondersignalfahren			4.89	.440 n.s.			8.78	.052 <sup>+</sup>
Homogen: durchschnittlich Spaß	2	3			0	0		
Homogen: viel Spaß	16	20			3	4		
Vollständig heterogen	16	15			9	1		
Heterogen: wenig und durchschnittlich in vergleichbaren Anteilen	8	3			1	1		
Heterogen: durchschnittlich und viel in vergleichbaren Anteilen	16	8			2	6		
Heterogen: wenig und viel in vergleichbaren Anteilen	7	6			1	2		

<sup>1</sup> n.s. = nicht signifikant; Chiquadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveaus: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001; <sup>2</sup> für Kreuztabellen werden ausschließlich vorhandene Ausprägungen angegeben; Erläuterung der Kategorien im Text

Tabelle 62: Prädiktion kurzfristigen und längerfristigen Erfolgs auf Resultateebene durch gruppenbezogene Voraussetzungen

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: Resultateebene, n ≤ 92	Kurzfristig nicht erfolgreich: Resultateebene, n ≤ 72	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Resultateebene, n ≤ 29	längerfristig nicht erfolgreich: Resultateebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Individuelle Bewertung der Atmosphäre in TG	M=4.88; SD = 0.388	M=4.89; SD = 0.371	-0.23	.969 n.s.	M=4.86; SD = 0.441	M=4.91; SD = 0.301	-0.13	.855 n.s.
Gruppenbewertung der Atmosphäre (M/Gruppe)	M=4.88; SD = 0.233	M=4.90; SD = 0.203	-0.35	.727 n.s.	M=4.88; SD = 0.268	M=4.90; SD = 0.226	-0.12	.890 n.s.
Homogenität dieser Bewertung (SD/Gruppe)	M=0.18; SD = 0.270	M=0.16; SD = 0.254	-0.36	.718 n.s.	M=0.14; SD = 0.239	M=0.14; SD = 0.228	-0.12	.890 n.s.
Individuelle Note für Training	M=1.82; SD = 0.967	M=1.67; SD = 0.635	-1.11	.268 n.s.	M=1.93; SD = 1.067	M=1.60; SD = 0.970	-1.56	.124 n.s.
Gruppennote (M/Gruppe)	M=1.78; SD = 0.816	M=1.75; SD = 0.671	-0.76	.451 n.s.	M=1.89; SD = 0.949	M=1.72; SD = 0.875	-1.29	.199 n.s.
Homogenität in der Note (SD/Gruppe)	M=0.54; SD = 0.213	M=0.57; SD = 0.249	-0.89	.373 n.s.	M=0.54; SD = 0.200	M=0.58; SD = 0.179	-2.03	.043*
Anzahl Trainingsteilnehmer			0.11	1.00 n.s.			1.67	.254 n.s.
sehr wenige (1-3)	3	3			0	0		
wenige (4-5)	50	38			14	14		
wie geplant (6)	39	31			15	7		
Zusammensetzung: Geschlecht			1.50	.710 n.s.			1.41	.634 n.s.
ausschließlich männlich	24	25			10	7		
überwiegend männlich	55	38			19	13		
ausgeglichen	5	3			0	1		
überwiegend weiblich	8	6			0	0		
Zusammensetzung: Alter			1.93	.628 n.s.			2.74	.140 n.s.
Homogen jung	2	4			0	0		
Homogen mittelalt	3	3			0	0		
Heterogen jede Altersgruppe	21	19			3	6		
Heterogen jung + mittelalt	66	46			26	15		
Zusammensetzung: Rettungsdienstenerfahrung			3.02	.760 n.s.			0.74	.488 n.s.
homogen Novizen	1	1			0	0		
homogen Intermediates	1	1			0	0		

Abschlussbericht „Achtung Blaulicht“

Merkmals <sup>2</sup>	Kurzfristig erfolgreich: Resultatebene, n ≤ 92	Kurzfristig nicht erfolgreich: Resultatebene, n ≤ 72	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: Resultatebene, n ≤ 29	längerfristig nicht erfolgreich: Resultatebene, n ≤ 21	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
heterogen alle Erfahrungsstufen	61	50			22	18		
heterogen Novizen + Intermediates	1	3			0	0		
heterogen Intermediates + Experten	6	2			0	0		
heterogen Novizen + Experten	22	15			7	3		
Zusammensetzung: Fahrerfahrung			2.55	.788 n.s.			1.40	.737 n.s.
homogen Intermediates	2	4			0	0		
Homogen Experten	16	13			4	5		
heterogen alle Erfahrungsstufen	33	20			13	7		
heterogen Novizen + Intermediates	2	1			0	0		
heterogen Intermediates + Experten	30	25			8	7		
heterogen Novizen + Experten	9	9			4	2		
Zusammensetzung: Spaß beim Sondersignalfahren			5.11	.410 n.s.			6.70	.254 n.s.
Homogen: durchschnittlich Spaß	5	6			2	0		
Homogen: viel Spaß	27	24			8	4		
Vollständig heterogen	24	22			11	6		
Heterogen: wenig und durchschnittlich in vergleichbaren Anteilen	10	4			3	3		
Heterogen: durchschnittlich und viel in vergleichbaren Anteilen	19	8			5	5		
Heterogen: wenig und viel in vergleichbaren Anteilen	7	8			0	3		

<sup>1</sup> n.s. = nicht signifikant; Chi-Quadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: \*p ≤ .10; \*\*p ≤ .05; \*\*\*p ≤ .01; \*\*\*\*p ≤ .001; <sup>2</sup> für Kreuztabellen werden ausschließlich vorhandene Ausprägungen angegeben; Erläuterung der Kategorien im Text

Tabelle 63: Prädiktion kurzfristigen und längerfristigen Gesamterfolgs durch gruppenbezogene Voraussetzungen

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: global, n ≤ 50	Kurzfristig nicht erfolgreich: global, n ≤ 82	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: global, n ≤ 13	längerfristig nicht erfolgreich: global, n ≤ 22	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
Individuelle Bewertung der Atmosphäre in TG	M=4.90; SD = 0.364	M=4.85; SD = 0.389	-0.92	.400 n.s.	M=5.00; SD = 0.000	M=4.73; SD = 0.551	-1.83	.134 n.s.
Gruppenbewertung der Atmosphäre (M/Gruppe)	M=4.86; SD = 0.293	M=4.90; SD = 0.190	-0.41	.684 n.s.	M=5.00; SD = 0.000	M=4.81; SD = 0.345	-2.42	.027*
Homogenität dieser Bewertung (SD/Gruppe)	M=0.15; SD = 0.280	M=0.16; SD = 0.244	-0.41	.689 n.s.	M=0.00; SD = 0.000	M=0.19; SD = 0.267	-2.42	.027*
Individuelle Note für Training	M=1.83; SD = 1.100	M=1.69; SD = 0.841	-0.45	.648 n.s.	M=1.42; SD = 0.494	M=2.00; SD = 1.310	-1.22	.236 n.s.
Gruppennote (M/Gruppe)	M=1.79; SD =1.007	M=1.76; SD = 0.652	-0.77	.446 n.s.	M=1.45; SD = 0.218	M=2.07; SD = 1.296	-1.34	.186 n.s.
Homogenität in der Note (SD/Gruppe)	M=0.51; SD = 0.091	M=0.56; SD = 0.253	-0.92	.400 n.s.	M=0.48; SD = 0.048	M=0.51; SD = 0.129	-0.02	.993 n.s.
Anzahl Trainingsteilnehmer			1.17	.597 n.s.			0.44	.727 n.s.
sehr wenige (1-3)	1	4			0	0		
wenige (4-5)	24	43			5	11		
wie geplant (6)	25	35			8	11		
Zusammensetzung: Geschlecht			6.92	.071 <sup>+</sup>			1.45	.655 n.s.
ausschließlich männlich	10	30			5	5		
überwiegend männlich	37	42			8	16		
ausgeglichen	1	5			0	1		
überwiegend weiblich	2	5			0	0		
Zusammensetzung: Alter			1.05	.832 n.s.			2.67	.274 n.s.
Homogen jung	1	4			0	0		
Homogen mittelalt	1	3			0	0		
Heterogen jede Altersgruppe	11	18			0	4		
Heterogen jung + mittelalt	37	57			13	18		
Zusammensetzung: Rettungsdienstenerfahrung			6.44	.269 n.s.			0.02	1.00 n.s.
homogen Novizen	0	2			0	0		
homogen Intermediates	1	1			0	0		

Abschlussbericht „Achtung Blaulicht“

Merkmal <sup>2</sup>	Kurzfristig erfolgreich: global, n ≤ 50	Kurzfristig nicht erfolgreich: global, n ≤ 82	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>	längerfristig erfolgreich: global, n ≤ 13	längerfristig nicht erfolgreich: global, n ≤ 22	χ <sup>2</sup> oder Z-Statistik	Unterschied <sup>1</sup>
heterogen alle Erfahrungsstufen	35	60			12	20		
heterogen Novizen + Intermediates	1	2			0	0		
heterogen Intermediates + Experten	6	2			0	0		
heterogen Novizen + Experten	7	15			1	2		
Zusammensetzung: Fahrerfahrung			8.71	.115 n.s.			1.99	.617 n.s.
homogen Intermediates	1	4			0	0		
Homogen Experten	12	12			3	5		
heterogen alle Erfahrungsstufen	21	22			7	10		
heterogen Novizen + Intermediates	0	3			0	0		
heterogen Intermediates + Experten	12	33			3	4		
heterogen Novizen + Experten	4	8			0	3		
Zusammensetzung: Spaß beim Sondersignalfahren			9.46	.091 <sup>+</sup>			8.87	.094 <sup>+</sup>
Homogen: durchschnittlich Spaß	2	6			1	0		
Homogen: viel Spaß	11	29			3	6		
Vollständig heterogen	13	20			7	4		
Heterogen: wenig und durchschnittlich in vergleichbaren Anteilen	9	3			0	2		
Heterogen: durchschnittlich und viel in vergleichbaren Anteilen	10	16			1	8		
Heterogen: wenig und viel in vergleichbaren Anteilen	5	8			1	2		

<sup>1</sup> n.s. = nicht signifikant; Chi-Quadrat- oder Mann-Whitney-U-Tests (exakte Testung, zweiseitig); Signifikanzniveau: <sup>+</sup>p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*\*p ≤ .001; <sup>2</sup> für Kreuztabellen werden ausschließlich vorhandene Ausprägungen angegeben; Erläuterung der Kategorien im Text

Für den Erfolg auf der Verhaltensebene sind nur wenige signifikante Unterschiede in gruppenbezogenen Merkmalen und hier vor allem im Bereich des längerfristigen Erfolgs zu finden (vgl. Tabelle 60). Personen, die längerfristig von den Trainings auf der Verhaltensebene profitieren, stammen aus Gruppen, die die Atmosphäre besser und homogener bewertet haben sowie aus Gruppen, die erneut hinsichtlich ihrer Freude am Fahren mit Sondersignal sehr heterogen waren und nicht nur aus mittel oder gerne fahrenden Teilnehmern zusammengesetzt waren. Für den Erfolg auf der Resultateebene gibt es ausschließlich einen Effekt im Bereich längerfristigen Erfolgs, längerfristig erfolgreiche Teilnehmer scheinen aus Gruppen zu stammen, die ihr Training homogener (aber - nicht signifikant – schlechter) bewertet haben.

Die Analysen zum Gesamterfolg (vgl. Tabelle 63) zeigen schließlich für den kurzfristigen Bereich einen Nachteil für Gruppen, die ausschließlich männlich oder ausgeglichen oder überwiegend weiblich besetzt waren sowie erneut den Unterschied hinsichtlich Freude am Fahren, dass erfolgreiche Personen von Gruppen mit vollständiger Heterogenität und Beteiligung von Personen mit wenig bzw. mittlerer Freude am Sondersignalfahren profitierten. Dieser Effekt findet sich auch für den längerfristigen Gesamterfolg. Zudem war längerfristiger Erfolg von Personen aus Trainingsgruppen bestimmt, die die Gruppenatmosphäre völlig übereinstimmend als sehr positiv bewerteten.

Bei allen Einteilungen nach Erfolg spielen Fahrerfahrung und Alterszusammensetzung erstaunlicherweise keine Rolle – deskriptiv scheinen aber heterogene Gruppen mit nicht zu weit auseinanderklaffenden Alters- und Erfahrungsstufen von Vorteil für den Trainingserfolg zu sein. Die Gruppengröße zeigte bei keiner Analyse systematische Unterschiede für den Trainingserfolg.

Um die Gruppenaspekte in ihrer Gesamtschau zu bewerten, werden im Folgenden jeweils die kurzfristig und langfristig erfolgreichste Trainingsgruppe beschrieben. Bei der kurzfristig erfolgreichsten Gruppe waren vier von fünf Teilnehmern auf der Lern- und Verhaltensebene erfolgreich und alle fünf auf der Resultateebene und im Gesamterfolg. Die Gruppe mit drei Männern und zwei Frauen bewertete die Trainingsatmosphäre unisono als sehr positiv und gab dem Training eine sehr gute Note. Alle Altersgruppen waren vertreten, alle hatten ihren Führerschein zehn Jahre oder länger und vier waren länger als 10 Jahre im Rettungsdienst tätig, ein Gruppenmitglied etwas kürzer. Drei der fünf Gruppenmitglieder gaben an, gar nicht und eher nicht gerne Sondersignal zu fahren, zwei fuhren eher durchschnittlich gern. Ein ähnliches Bild ergibt sich für die längerfristig erfolgreichste Gruppe. Diese Trainingsgruppe bestand aus sechs Personen, von denen vier an der T4-Messung teilnahmen. Von diesen vier waren zwei auf der Lernebene erfolgreich, drei auf der Verhaltensebene und im Gesamterfolg sowie alle vier auf der Resultateebene. Bewertung des Trainings und der Atmosphäre waren nahezu identisch mit der kurzfristig erfolgreichen Gruppe. Von den fünf Männern und einer Frau waren vier Personen unter 30 Jahre alt, zwei waren zwischen 31 und 50 Jahren. Fünf hatten seit über zehn Jahren ihren Führerschein, eine Person knapp unter zehn Jahren, wobei die Erfahrung im Rettungsdienst ebenso wie die Freude am Fahren mit Sondersignal vollständig heterogen war. Interessant hervorzuheben ist neben der wichtigen Rolle heterogener Zusammensetzungen, dass beide „erfolgreichsten“ Gruppen mit erfahrenen Autofahrern besetzt waren - auch wenn sich für diese Variable keine statistischen Effekte auf den Erfolg gezeigt haben, scheint dies zusammen mit den leichten individuellen Effekten in diesem Bereich (vgl. Abschnitt 6.3) der häufiger auch in den qualitativen Trainingsevaluationen geäußerten Annahme zu widersprechen, dass vor allem unerfahrenere Fahrer von dem Training profitieren würden.

Wird die Betrachtungsebene von systematischen Gruppenunterschieden auf individuelle Veränderungen und Erfolg verschoben, lassen sich die Ergebnisse wie folgt zusammenfassen:

- Vor dem Hintergrund, dass Personen sich nicht nur in einzelnen Aspekten, sondern auch in Mustern positiver und negativer Entwicklungen ändern können, wurden neben Einzelindikatoren auch ebenspezifische und globale Erfolgsindikatoren entwickelt und untersucht, ob sich Zusammenhänge zu individuellen, organisatorischen und gruppenspezifischen Merkmalen finden lassen.
- Die Anteile erfolgreicher Personen variieren stark zwischen den verschiedenen Erfolgsindikatoren. Bei dem „härtesten“ globalen Indikator sind zwischen 37% und 38% der Teilnehmer erfolgreich, bei den verschiedenen ebenspezifischen Indikatoren liegen die Anteile zwischen 54% und 59%.
- Insgesamt finden sich vergleichsweise wenige Faktoren, die prädiktiv für die verschiedenen Erfolgsfaktoren sind.
- Am ehesten zeigen sich Einflüsse individueller Rettungsdiensterfahrung und der Freude am Fahren mit Sondersignal – hier profitieren eher diejenigen, die weniger gerne fahren – sowie Charakteristika der Trainingsgruppen: Mitglieder heterogen zusammengesetzter Gruppen und Gruppen, die nicht nur „begeisterte“ Fahrer enthalten, profitieren eher von den Trainings als andere.



## 7. Einflussfaktoren auf Sondersignalfahrten im Rettungsdienst

Viele Faktoren können eine Einsatzfahrt beeinflussen. Witterungsbedingungen verändern oft die Fahrweisen, wenn das eigene Unfallrisiko beispielsweise anders eingeschätzt wird. Zusätzlich spielt der Faktor Mensch natürlich die größte Rolle für die Fahrweise - neben personennahen Faktoren und der allgemeinen Tagesverfassung können auch kurzzeitig externe Faktoren relevant sein. Ein großer Einflussfaktor, der häufig im Rettungsdienst genannt wird, ist das Einsatzstichwort, aber auch die Schichtlänge kann sich auf das Fahren auswirken. In diesem Abschnitt werden diese potenziellen Einflüsse auf die Fahrweise untersucht.

Zunächst werden Fahrten zum Einsatzort und dann mit Patient in die Klinik verglichen, anschließend werden die Einflüsse verschiedener Schichtaspekte auf Einsatzfahrten betrachtet und abschließend die Sichtungen der Fahrvideos vorgestellt und ein Überblick gegeben, wie oft es bei Einsatzfahrten zu potenziell kritischen Situationen kommt und wie diese aussehen.

### 7.1 Einsatz-Hinfahrten im Vergleich zu Patiententransporten

Tabelle 64 gibt einen Überblick über die Mittelwerte und Standardabweichungen der Fahrprofilaten zum ersten Messzeitpunkt zwischen den verschiedenen Arten der Fahrten und zeigt an, ob sich signifikante Unterschiede zeigen.

Tabelle 64: Unterschiede zwischen RTW- Sondersignalfahrten zum Einsatzort und Sondersignalfahrten mit Patient

Abhängige Variable	RTW Fahrten zum Einsatzort (N=121)		RTW Patienten-transporte zur Klinik (N=121)		Mittelwert-vergleiche	
	M	SD	M	SD	T	p
Durchschnittliche Geschwindigkeit (in km/h)	53.1	10.9	53.4	16.1	-0.27	.785
Maximale Geschwindigkeit (in km/h)	117.5	20.6	107.4	28.7	4.44	.000***
Maximale positive Längsbeschleunigung	3.57	2.01	1.96	1.60	11.55	.000***
Mittlere positive Längsbeschleunigung	0.32	0.14	0.23	0.14	13.75	.000***
Maximale negative Längsbeschleunigung	-3.51	1.59	-2.22	1.39	-10.67	.000***
Mittlere negative Längsbeschleunigung	-0.33	0.14	-0.24	0.15	-13.40	.000***
Mittlere positive Querb beschleunigung	0.48	0.28	0.39	0.33	8.06	.000***
Mittlere negative Querb beschleunigung	-0.45	0.27	-0.37	0.32	-8.26	.000***

Berechnung: T-Test für verbundene Stichproben; Abkürzungen: N=Anzahl der Probanden, M=Mittelwert, SD=Standardabweichung, T=Prüfgröße des T-Tests, p=Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; alle Beschleunigungen in m/s<sup>2</sup>

Hinsichtlich der Fahrprofilaten werden jeweils die maximale und mittlere Geschwindigkeit, Beschleunigung und Bremsverzögerung sowie die mittlere Querb beschleunigung ausgewertet. Da die Querb beschleunigung insgesamt sehr stark zwischen verschiedenen Geschwindigkeiten schwankt, wird dabei nur die mittlere Querb beschleunigung berichtet. Zur Ermittlung der Unterschiede werden jeweils nur die Probanden herangezogen, die sowohl Fahrten zum Einsatzort als auch Patiententransporte gefahren sind. Damit entfallen 20 Probanden, die nur Hinfahrten gemacht haben, Patienten aber nicht unter Nutzung des Sondersignals in die Klinik gebracht haben.

Die Fahrten zum und mit Patienten unterscheiden sich in allen betrachteten Variablen mit der Ausnahme der mittleren Geschwindigkeit hochsignifikant. Das zeigt, dass die Probanden mit Patienten deutlich ruhiger

fahren, aber auch, dass die durchschnittliche Geschwindigkeit dabei nicht geringer ist. Auch muss mit Patient seltener sehr stark gebremst werden, was zum einen auf die höhere maximale Geschwindigkeit ohne Patient zurückzuführen sein kann. Es könnte jedoch auch sein, dass die Probanden mit Patienten etwas vorausschauender fahren und somit seltener in kritische Situationen kommen.

Diese Unterschiede sind genauso auch zu den Zeitpunkten T2 und T3 in den einzelnen Versuchsgruppen zu finden; die mittlere Geschwindigkeit unterscheidet sich jeweils nicht zwischen den Hinfahrten zum Einsatzort und den Patiententransporten in die Klinik, jedoch die maximale Geschwindigkeit und die Beschleunigungen.

Zum Zeitpunkt T4, zu dem nur noch 26 Probanden sowohl mit als auch ohne Patient mit Sondersignal gefahren sind, unterscheidet sich auch die maximale Geschwindigkeit nicht mehr signifikant (Hinfahrten:  $M=112.7$  km/h,  $SD= 22.0$  km/h; Patiententransporte:  $M=106.7$  km/h,  $SD=32.0$ ;  $T=1.12$ ,  $p=.270$ ). Alle Beschleunigungsvariablen bleiben jedoch selbst bei der kleinen Stichprobe signifikant höher, wenn kein Patient transportiert wird im Vergleich zu dem Patiententransport.

Die Unterschiede zwischen den Fahrten zum Einsatzort und zur Klinik zeigen, dass die Einsatzfahrten zum Einsatzort deutlich rasanter durchgeführt werden. Die Fahrten mit Patient im Fahrzeug zeigen, dass die Einsatzfahrer auch unter Nutzung der Sondersignale anders fahren können.

## 7.2 Einfluss von Schichtvariablen auf Sondersignalfahrten

In diesem Abschnitt soll überprüft werden, welche Variablen, die über die Schicht hinweg gemessen wurden, einen Einfluss auf die Fahrweise haben. Da die Sondersignalfahrten pro Messzeitpunkt aus verschiedenen Schichten kommen, werden jeweils nur die Schichten in die Berechnung einbezogen, für die auch Fahrdaten vorliegen. Da die Anzahl der Schichten über die Probanden hinweg unterschiedlich groß ist, werden jeweils die Mittelwerte der Variablen genutzt.

Folgende Variablen werden in die Berechnungen einbezogen:

- Die mittlere Schwere der Einsatzstichworte jeweils getrennt hinsichtlich der Einsatzfahrten zum Einsatzort und mit Patient in die Klinik. Der mögliche Bereich liegt zwischen 1 (geringe Schwere) und 3 (hohe Schwere).
- Die mittlere Art der Schicht setzt sich zusammen aus den Arten der Schichten, die gemessen wurden mit folgender Kodierung: 1=Frühschicht, 2=Spätschicht, 3=Tagschicht, 4=Nachtschicht.
- Die mittlere Schichtdauer, d.h. Schichtlänge bereinigt um Überstunden und den späteren Einbau der Messgeräte.
- Die mittlere Anzahl der Einsätze, die über die gemessenen Schichten abgearbeitet werden mussten. Dort hinein fallen alle Einsätze, egal ob mit oder ohne Sondersignal zum Einsatzort gefahren wurde und ob bzw. wie der Patient transportiert wurde.
- Die durchschnittliche Anzahl der selbstberichteten kritischen Ereignisse während der Einsatzfahrten über die Schichten.
- Die mittlere Müdigkeit und Ermüdung über die Schichten.
- Die mittlere geistige und zeitliche Anforderung über die Schichten.

Tabelle 65 gibt einen Überblick über die Korrelationen zwischen den abhängigen Variablen und diesen schichtbezogenen potenziellen Einflussfaktoren.

Tabelle 65: Korrelationen zwischen den abhängigen Variablen der RTW-Sondersignalfahrten und schichtbezogenen Kontrollvariablen

Einsatzfahrten/ Kontrollvariablen	1 <sup>1</sup>	2 <sup>1</sup>	3 <sup>1</sup>	4 <sup>1</sup>	5 <sup>1</sup>	6 <sup>1</sup>	7 <sup>1</sup>	8 <sup>1</sup>	9 <sup>1</sup>
<b>Hinfahrten</b>									
a <sup>1</sup>	.276***	.189*			-.172 <sup>+</sup>	-.471***			
b <sup>1</sup>	.157 <sup>+</sup>		-.166	-.218*	-.188*	-.179*			.217*
c <sup>1</sup>		-.271**	-.313***	-.273**			.164 <sup>+</sup>	.204*	
d <sup>1</sup>		-.255**	-.265**	-.217*	-.181 <sup>+</sup>	-.149 <sup>+</sup>			.214*
e <sup>2</sup>		.144 <sup>+</sup>	.218**			-.174*	-.231*		
f <sup>2</sup>		.236**	.291***	.209*	.204*				-.219*
g <sup>1</sup>		-.194	-.242**	-.194*	-.196*				.257**
h <sup>2</sup>		.184	.247**	.169*	.200*				-.258**
<b>Patiententransporte</b>									
a <sup>1</sup>		.239**	.171 <sup>+</sup>			-.339***			
b <sup>1</sup>					-.198 <sup>+</sup>	-.303***			
c <sup>1</sup>	-.182*		-.207*	-.229*			.259**	.166 <sup>+</sup>	
d <sup>1</sup>			-.195*	-.271**					
e <sup>2</sup>			.236**	.184*					
f <sup>2</sup>		.165 <sup>+</sup>	.263**	.248**	.215*				
g <sup>1</sup>		-.173 <sup>+</sup>	-.242**	-.295**	-.215*		.196*	.171 <sup>+</sup>	
h <sup>2</sup>		.168 <sup>+</sup>	.234**	.260**	.209*		-.198*	-.161 <sup>+</sup>	

Einsatzfahrten (Zeilen): a=Durchschnittliche Geschwindigkeit (in km/h), b=Maximale Geschwindigkeit (in km/h), c=Maximale positive Längs-beschleunigung (in m/s<sup>2</sup>), d=Mittlere positive Längs-beschleunigung (in m/s<sup>2</sup>), e=Maximale negative Längsbeschleunigung (in m/s<sup>2</sup>), f=Mittlere negative Längsbeschleunigung (in m/s<sup>2</sup>), g=Mittlere positive Quer-beschleunigung (in m/s<sup>2</sup>), h=Mittlere negative Quer-beschleunigung (in m/s<sup>2</sup>)

Kontrollvariablen (Spalten): 1=mittlere Einsatzschwere (getrennt nach Hinfahrten und Patiententransporten), 2=mittlere Schichtdauer, 3=mittlere Schichtart, 4=mittlere Müdigkeit während der Schicht, 5=mittlere Ermüdung über die Schicht, 6=mittlere Anzahl der Einsätze, 7=mittlere geistige Anforderungen während der Schicht, 8=zeitliche Anforderungen während der Schicht, 9=mittlere Anzahl selbstberichteter kritischer Ereignisse

Berechnung: Korrelation nach Pearson, Signifikanzniveaus: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001; nur Korrelationen auf dem 10%-Signifikanzniveau werden berichtet

<sup>1</sup> positive Ausprägungsrichtung: je höher der Wert, desto höher die Ausprägung der Variable

<sup>2</sup> negative Ausprägungsrichtung: je niedriger der Wert, desto höher die Ausprägung der Variable

Die *Einsatzschwere*, von der vermutet wird, dass sie sich auf die Fahrweise der Einsatzfahrer auswirkt, zeigt auf den Hinfahrten zum Einsatzort tatsächlich einen positiven Zusammenhang zur mittleren und maximalen Geschwindigkeit. Je höher die Schwere der Einsätze, desto schneller fahren die Probanden. Jedoch gibt es keine Zusammenhänge zwischen der Schwere und den Beschleunigungen. Bei den Patiententransporten zeigt sich der Zusammenhang mit der Geschwindigkeit nicht. Dort findet sich jedoch ein negativer Zusammenhang zur maximalen Beschleunigung. Je schwerer die Verdachtsdiagnose war, desto weniger stark wurde auf der Fahrt in die Klinik beschleunigt. Dieser Zusammenhang könnte verdeutlichen, dass Einsatzfahrer bei schwerer verletzten/erkrankten Personen versuchen, im Sinne des Wohlbefindens des Patienten nicht zu stark zu beschleunigen. Im Mittelwert der Beschleunigung ist der Effekt nicht zu finden.

Die *Schichtdauer* weist mit den meisten Fahrprofilaten Zusammenhänge auf. Je länger die Schicht ist, desto höher sind die gefahrenen mittleren Geschwindigkeiten, jedoch nicht die maximalen. Längere Schichtdauern finden sich häufig in ländlicheren Wachen und seltener in den großen Städten. Die höheren mittleren Geschwindigkeiten können daher auf häufigere Fahrten außerorts zurückgehen. Die höheren Beschleunigungen finden sich eher bei den Probanden mit kürzeren Schichtzeiten, also eher bei den städtischen Rettungsdienstfahrern. Die maximale Geschwindigkeit hat jedoch keinen Zusammenhang zur Schichtdauer.

Die *Schichtart* verhält sich hinsichtlich der Zusammenhänge zu den Fahrprofilen ähnlich der Schichtdauer. Eine Erklärung der Zusammenhänge könnten hier die Nachtschichten bieten, in denen häufig weniger Verkehr auftritt und daher weniger häufig stark beschleunigt oder gebremst wird. Die maximalen Geschwindigkeiten nehmen mit Höhe der Schichtart ab, also zeigen sich in Früh- oder Spätschichten eher höhere maximale Geschwindigkeiten als zu Nachtschichten. Bei Patiententransporten hingegen nimmt die mittlere Geschwindigkeit in späteren Schichten zu.

Die *Müdigkeit und Ermüdung* zeigen ähnliche Zusammenhänge zu den Fahrprofilen. Je höher die Müdigkeit oder Ermüdung ist, desto langsamer wurde in den Schichten gefahren (sowohl im Mittel als auch im Maximum) und desto weniger wurde beschleunigt. Da es sich hierbei nur um eine Korrelation handelt, kann keine Aussage über die Richtung eines Einflusses gemacht werden. Inhaltlich ergibt es jedoch Sinn, dass Einsatzfahrer, die müder sind, eher ruhiger fahren, um hinsichtlich einer Anpassung der Risikowahrnehmung ihr eigenes Unfallrisiko zu verringern.

Die *Anzahl der Einsätze*, die während der Schichten insgesamt abgearbeitet wurden, korrelieren bei Hinfahrten und Patiententransporten negativ mit der gefahrenen Geschwindigkeit und der mittleren Beschleunigung bei Hinfahrten. Je mehr Einsätze die Probanden während der Schicht hatten, desto langsamer sind sie gefahren. Hier könnte einerseits wieder ein Gefälle zwischen eher städtischen und eher ländlichen Rettungswachen zugrunde liegen. In einigen Wachen sind die Probanden zwischen den Einsätzen nicht zurück zur Wache gekommen, da sie von Einsatz zu Einsatz gefahren sind. Die Schichtlänge und Anzahl der Einsätze korrelieren nicht miteinander, dafür aber die Anzahl der Einsätze mit der Einsatzschwere. Je mehr Einsätze gefahren werden, desto geringer ist die mittlere Einsatzschwere ( $r=-.217^{**}$ ). Es werden also viele leichtere Einsätze oder weniger schwere gefahren. Der Zusammenhang zur Geschwindigkeit könnte darüber vermittelt sein.

Die *geistigen und zeitlichen Anforderungen* korrelieren positiv mit der maximalen Beschleunigung und bei den Patiententransporten auch mit der Höhe der Querschleunigungen. Je stärker beschleunigt wird, desto eher werden geistige und zeitliche Anforderungen auch hoch bewertet. Untereinander korrelieren die Anforderungen nicht, auch nicht mit den anderen Kontrollvariablen, bis auf die Anzahl der Einsätze, die positiv mit der zeitlichen Anforderung korrelieren ( $r=.274^{***}$ ).

Die *Anzahl der selbstberichteten kritischen Ereignisse* zeigt lediglich bei den Hinfahrten Zusammenhänge zu den Fahrprofilaten. Je mehr kritische Ereignisse angegeben werden, desto höher die mittleren Beschleunigungen. Es kann sein, dass durch viele kritische Ereignisse häufiger abgebremst werden muss und somit auch wieder beschleunigt oder dass aufgrund vieler stärkerer Beschleunigungen auch häufiger kritische Ereignisse geschehen.

Neben den reinen Korrelationsanalysen wurden auch durch vorwärtsselektierende Regressionsanalysen untersucht, welche Variablen die jeweils größten Einflüsse auf die Fahrprofile haben. Im ersten Schritt wurde jeweils das Alter und Geschlecht kontrolliert. Im zweiten Schritt wurden erfahrungsspezifische Variablen (Qualifikation, Rettungsdienst erfahrung, Fahrzeiten im Rettungsdienst, Anzahl der Führerscheinklassen und die Verkehrserfahrung) schrittweise hinzugefügt, gefolgt von den Schichtdaten (Einsatzschwere, Schichtdauer, Schichtart, Müdigkeit, Ermüdung, geistige Anforderungen, zeitliche Anforderungen und Anzahl kritischer Ereignisse).

Ein Erfahrungseinfluss findet sich lediglich bei der mittleren Geschwindigkeit der Fahrten zum Einsatzort. Personen, die mehr Führerscheinklassen besitzen, zeigen höhere mittlere Geschwindigkeiten. Alle anderen Erfahrungsvariablen zeigen keine signifikanten Einflüsse auf die Fahrprofile. Hinsichtlich der schichtbezogenen Variablen gibt es unterschiedliche Einflussfaktoren, am häufigsten (insgesamt bei 12 der 16 Analysen) zeigt die Schichtart einen signifikanten Einfluss auf das Fahrprofil. Je später die gemessenen Schichten lagen, desto geringer wurde in alle Richtungen sowohl im Mittel als auch maximal beschleunigt und desto geringer war die maximale Geschwindigkeit bei Patiententransporten. Bei acht Analysen zeigt sich ein Einfluss der Anzahl der Einsätze. Je mehr Einsätze gefahren werden, desto geringer sind die Beschleunigungen (mittlere Quer- und Längsbeschleunigung) und mittleren Geschwindigkeiten. Die maximalen Geschwindigkeiten bei Hinfahrten zum Einsatzort steigen allerdings mit einer höheren Zahl an Einsätzen. In sieben Fällen zeigt sich ein signifikanter Einfluss der geistigen Anforderungen. Werden die geistigen Anforderungen höher bewertet, fahren die Probanden rasanter und mit höheren Beschleunigungen. Höhere Ermüdung über die Schicht führte dahingegen bei Patiententransporten in drei Beschleunigungsmaßen zu geringeren Beschleunigungen. Die höheren empfundenen zeitlichen Anforderungen bewirken zudem eine Erhöhung der mittleren Geschwindigkeit bei Einsatzfahrten zum Einsatzort. Alle Regressionsmodelle mit den genauen Ergebnissen befinden sich im Anhang A.

## 7.3 Videoanalysen

Insgesamt konnten sehr viele Videos über die Schichten und Zeitpunkte hinweg gesammelt werden. Wie bereits weiter oben beschrieben waren diese besonders wichtig, um die Fahrprofilaten überhaupt hinsichtlich der Einsatzfahrten sortieren zu können. Zusätzlich können jedoch auch weitere Informationen aus den Videos gewonnen werden. Die Zahlen, wie oft es während einer Einsatzfahrt zu einem kritischen Ereignis kommt, wurde mithilfe von 54 Minuten Videodaten ermittelt (Unterkofler & Schmiedel, 1994) und ist bereits etwas älter. Da die Videosichtung insgesamt sehr viel Zeit in Anspruch nimmt, wurde in der vorliegenden Studie eine Zufallsauswahl an Videos gezogen und gesichtet. Zuvor wurde mittels drei verschiedener Sichter ein System zur Bewertung der Videos entwickelt. Dieses war angelehnt an die Studie der BAST (Unterkofler & Schmiedel, 1994) und soll im Folgenden zunächst vorgestellt werden.

### 7.3.1 Entwicklung des Analyseverfahrens zur Videosichtung

Um kritische Fahrsituationen während der Einsatzfahrten mit Sondersignal ermitteln zu können, bedarf es eines möglichst eindeutigen Beobachtungssystems der Situationen und einer Festlegung, was als kritisches Ereignis zählt. Angelehnt an die kritischen Fahrsituationen der BAST-Studie (Unterkofler & Schmiedel, 1994) und in einem mehrstufigen Prozess konnte ein Beobachtungssystem entwickelt werden. Mithilfe von drei Beurteilern wurde das System entwickelt und getestet.

Folgendes System wurde abschließend entwickelt. Ereignisse wurden zunächst hinsichtlich der Kategorie bewertet (Vorfahrtregelung, Fußgängerüberwege, Überholvorgänge und sonstige Situationen, in denen ein Eingreifen erforderlich ist). Fährt ein Einsatzfahrer beispielsweise über eine rote Kreuzung, wird die Vorfahrtregelung kodiert, überholt er auf einer Straße unabhängig einer Vorfahrtssituation, wird Überholen kodiert. Überholvorgänge direkt vor Kreuzungen gehören zum Ereignis Vorfahrtregelung. Sobald ein Ereignis detektiert wurde und die Kategorie festgelegt war, wurden alle Unterkategorien des Ereignisses bewertet.

- Straßentyp: Bewertung, ob das Ereignis in einer Fußgängerzone, auf städtischen Straßen, überörtlichen Straßen oder Autobahnen stattgefunden hat.
- Art des Ereignisses: Bewertung, was für eine Kreuzung (bspw. rot, Vorfahrt gewähren oder Kreisverkehr) überfahren wurde, ob es sich bei den Fußgängerüberwegen um eine Ampel oder

einen beschilderten Weg handelt, was für eine Art Überholen vorlag (bspw. unübersichtliche Gerade oder übersichtliche Kurve) oder welche sonstige Art beobachtet wurde (bspw. Tiere, Fahrt gegen die Fahrtrichtung oder Engstellen)

- Größe: die Anzahl der Fahrsteifen vor dem Ereignis wurden kodiert (einer, zwei oder mehr)
- Verkehrsdichte: die Anzahl der Fahrzeuge vor und während des Ereignisses in Fahrtrichtung
- Querverkehr: Fahrzeuge und Fußgänger die jeweils von rechts oder links hinsichtlich der Fahrtrichtung des Einsatzfahrers kamen
- Gegenverkehr: Fahrzeuge, die dem Einsatzfahrzeug entgegen kamen
- Verkehr in Fahrtrichtung (beim Überholen): Kodierung, ob die zu überholenden Fahrzeuge stehen oder selbst fahren

Es wurden jeweils alle relevanten und möglichen Unterkategorien bewertet. Sie sind für die einzelnen Kategorien teilweise etwas unterschiedlich, können also zum Teil nur innerhalb der Kategorien verglichen werden. Unterkategorien, die nicht relevant waren, wurden frei gelassen.

Zusätzlich wurde in jedem erfassten Ereignis die Witterung erfasst mithilfe der Bodenverhältnisse, Luftverhältnisse und Lichtverhältnisse. Auch die Reaktion auf das Ereignis wurde jeweils erfasst.

Die detektierten Situationen sind nicht automatisch kritische Situationen, so kann es beispielsweise sein, dass ein Einsatzfahrer auf einer übersichtlichen Gerade ein langsam fahrendes Fahrzeug überholt, welches das Einsatzfahrzeug wahrgenommen hat. Oder es wird über eine gelbe Ampel gefahren, die in alle Richtungen überschaubar ist und in deren Nähe keinerlei andere Fahrzeuge oder Fußgänger sind. Dennoch erfordern alle detektierten Ereignisse eine erhöhte Aufmerksamkeit des Einsatzfahrers, denn entweder werden Verkehrsregeln überschritten (bspw. das Überfahren einer roten Ampel), ausgereizt (überholen, wo es sonst eigentlich unüblich ist) oder es ist eine Reaktion erforderlich (z.B. Anpassung der Fahrtlinie, um sich durch stockenden Verkehr zu manövrieren). Daher sollen alle Ereignisse, die in den Videos ermittelt werden, als potenziell kritische Situationen bezeichnet werden.

Zur Erstellung des Beobachtungssystems wurden insgesamt 212 Minuten Einsatzfahrten genutzt. Die Übereinstimmung der Detektion der potenziell kritischen Ereignisse zwischen allen drei Ratern lag bei neun Einsatzfahrten über insgesamt 62 Minuten bei ICC=.615, die Klassifikation nach Ereigniskategorie bei ICC=.936. Am häufigsten führten unterschiedlich lange Beobachtungszeiträume (Beginn bis Ende des Ereignisses) zu Unstimmigkeiten in der Übereinstimmung. Nach der letzten Überarbeitung der Beobachtungsbögen wurde ein vierter Rater in die Bewertung eingearbeitet und getestet. Zur Einarbeitung wurden einige Fahrten genutzt, die bereits vorher bewertet wurden. Da das System zwischendurch immer wieder angepasst werden musste, können die bis dahin gesichteten Einsatzfahrten nicht in die Auswertung einbezogen werden. Einige Fahrten wurden zur Testung des Systems auch explizit hinsichtlich der Witterung und Sichtverhältnisse herausgesucht (Schnee, Windböen, nächtliche Fahrten) und würden daher die Auswertungen verfälschen, wenn die aufgenommen werden.

Daher wurde insgesamt eine Zufallsziehung der Einsatzfahrten vorgenommen. Alle Fahrten auf dem Weg zum Einsatzort unter Nutzung der Sonder- und Wegerechte zwischen 4 und 10 Minuten wurden in die Ziehung aufgenommen und pro Proband und Messzeitpunkt wurde jeweils ein Video gezogen. Gab es keine Videos mit dem Zeitfenster, wurden entweder zwei kürzere oder ein längeres gezogen.

So wurden insgesamt 448 Videos zufällig gezogen. Die Gesamtdauer von 2986 Minuten dieser Videos war jedoch noch immer zu groß, um sie alle bewerten zu können. Nach Ausschluss der Probanden, die aus der Studie zurückgetreten sind oder ausgeschlossen werden mussten, blieben noch 375 Videos mit insgesamt 2512 Minuten Einsatzfahrten. Davon wurden insgesamt 172 Videos mit insgesamt 1138 Minuten gesichtet. Dieser Wert ist bereits deutlich höher als bisherige Videosichtungen der BAST-Studie (Unterkofler &

Schmiedel, 1994) und sollte daher eine gute Möglichkeit zur Erneuerung der Zahlen für potenziell kritische Ereignisse liefern können.

### 7.3.2 Potenziell kritische Situationen bei Sondersignalfahrten

Im Folgenden werden die Daten für die Messungen vor dem Training dargestellt (T1 für Gruppe 1 sowie T1 und T2 für Gruppe 2), um einen Überblick zu bekommen, wie die Verteilung der potenziell kritischen Ereignisse aussieht. Insgesamt 54 Probanden mit ca. 520 Minuten Einsatzfahrten zum Einsatzort wurden für den Messzeitpunkt T1 gesichtet. Dabei waren 33 Probanden in der Gruppe 1 und 21 in der Gruppe 2 (mit je zwei Messzeitpunkten).

In den 520 Minuten wurden 845 potenziell kritische Situationen während der Fahrten ermittelt. Das ergibt 1.63 potenziell kritische Situation pro Minute oder alle 36.9 Sekunden eine potenziell kritische Situation. Hierbei sind Situationen innerorts und außerorts zusammengenommen, jedoch überwiegen die Anteile innerorts. Neun der 54 Probanden sind mindestens einmal mit einem NEF gefahren (insgesamt 90 Minuten), alle anderen mit RTW (insgesamt 430 Minuten), somit überwiegen die Anteile der RTW-Fahrten. Werden die Ereignisse hinsichtlich des Fahrzeugs aufgeteilt betrachtet, ergeben sich für RTW alle 39.1 Sekunden und für NEF alle 29.0 eine potenziell kritische Situation. Viele der Situationen sind in der Wahrnehmung der Einsatzfahrer jedoch ganz normale Fahrsituationen, weshalb an dieser Stelle noch einmal deutlich gemacht werden soll, dass die Situationen nicht zwangsweise kritisch sein müssen, aber eine erhöhte Aufmerksamkeit erfordern.

Interessant ist jedoch nicht nur die Anzahl der potenziell kritischen Ereignisse, sondern auch, wie lange diese andauern. Eine rote Kreuzung zu überfahren kann beispielsweise sehr schnell gehen, wenn alles überschaubar ist und keine Fahrzeuge da sind, kann sich aber auch lange hinziehen, wenn der Einsatzfahrer nicht zwischen den anderen Fahrzeugen hindurch kommt. Eine Überholssituation kann lange dauern, wenn eine lange Kolonne am Stück überholt wird oder nur kurz, wenn ein fast stehendes Fahrzeug ohne Gegenverkehr direkt überholt werden kann. Je länger eine Situation dauert, desto länger müssen die Einsatzfahrer sich auf die potenziell kritische Situation konzentrieren. Im Schnitt dauern die potenziell kritischen Situationen vor dem Training 6 Sekunden mit einem Bereich zwischen einer und 44 Sekunden. Insgesamt waren ca. 87 der 520 gefahrenen Minuten während potenziell kritischer Situationen. In 16.7% der Einsatzfahrten mussten die Einsatzfahrer demzufolge besondere Aufmerksamkeit walten lassen.

**Kategorie der potenziell kritischen Situationen**

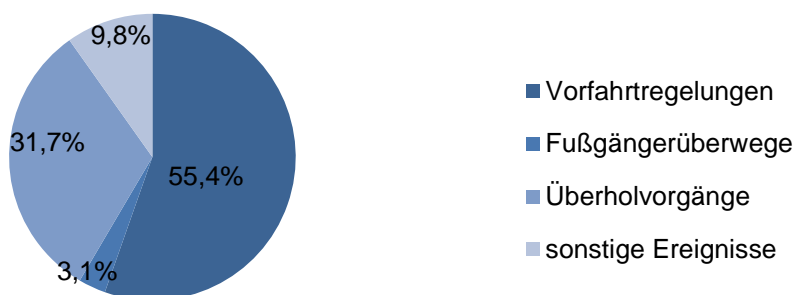


Abbildung 50: Verteilung der Kategorien potenziell kritischer Ereignisse vor dem Training

Im Folgenden sollen die Situationen genauer dargestellt werden. Zunächst werden die Variablen, die für alle Kategorien gleich sind, dargestellt (beispielweise Ortslage und Witterung). Daran anschließend werden die vier Kategorien einzeln beschrieben hinsichtlich der spezifischen Bewertungen (z.B. Verkehrsdichte, Art des Ereignisses, andere Verkehrsteilnehmer).

Über die Hälfte der potenziell kritischen Situationen fallen auf Vorfahrtregelungen (55.4%), gefolgt von Überholvorgängen (31.7%). Nur 3.1% der Ereignisse finden an Fußgängerampeln oder -überwegen statt. 9.8% fallen unter sonstige potenziell kritische Situationen.

Die meisten Ereignisse finden auf städtischen Straßen statt, gefolgt von überörtlichen Straßen (bis max. 100 km/h Höchstgeschwindigkeit). Die Kategorie Überholen findet im Vergleich zu den anderen Kategorien prozentual am häufigsten auch außerorts (besonders auf Landstraßen, aber auch auf Autobahnen) statt. Fußgängerzonen tauchen in der Stichprobe vor dem Training nicht auf. Die Vorfahrtssituationen auf Autobahnen betreffen das Auf- und Abfahren der Autobahn (vgl. Abbildung 51).

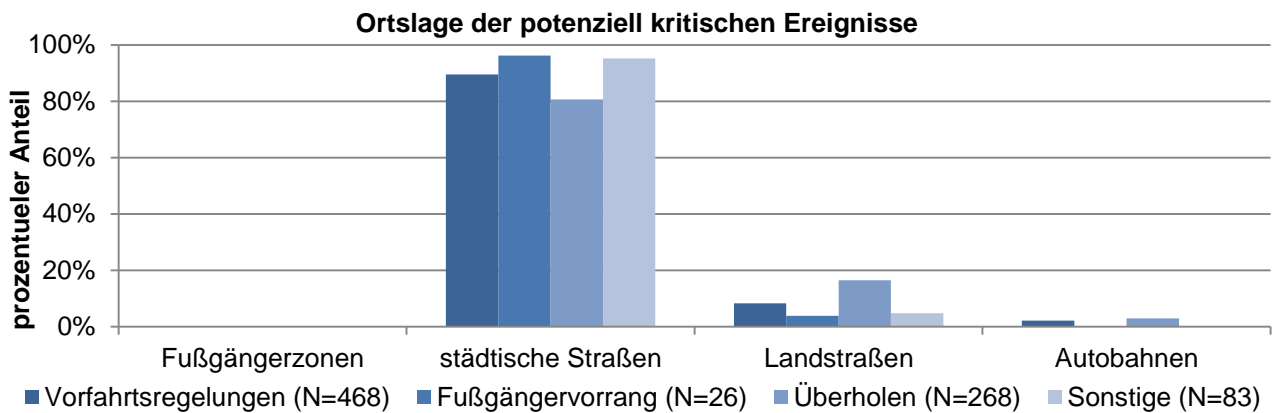


Abbildung 51: Ortslagen der potenziell kritischen Ereignisse in Abhängigkeit der Ereigniskategorien

Die meisten Straßen haben zum Zeitpunkt des potenziell kritischen Ereignisses nur eine Spur, sehr selten gibt es mehr als zwei Spuren in Fahrtrichtung. Dieses Bild ist zu erwarten, da die meisten Ereignisse innerorts kodiert wurden. Überholvorgänge können teilweise durch zwei Spuren verringert werden, da nicht in den Gegenverkehr ausgewichen werden muss, wenn die übrigen Verkehrsteilnehmer das Einsatzfahrzeug wahrnehmen und die rechte Spur benutzen (vgl. Abbildung 52).

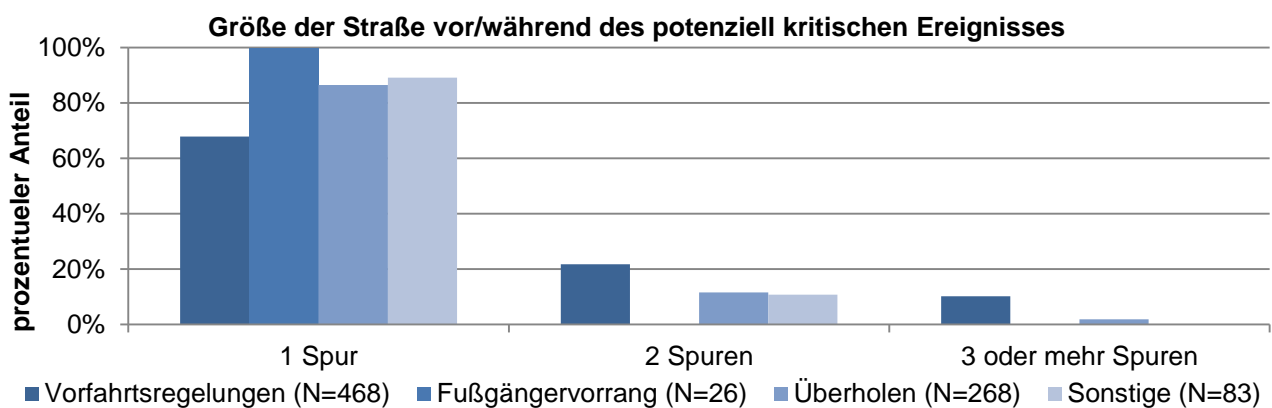


Abbildung 52: Straßengröße in Fahrtrichtung vor oder während des potenziell kritischen Ereignisses in Abhängigkeit der Ereigniskategorie vor dem Training

Hinsichtlich der Witterung hat sich innerhalb der betrachteten Videos über die Zeit nur in vier Videos etwas geändert. Einmal hat während der Fahrt leichter Regen eingesetzt, was zu einer feuchten Straße führte. Einmal war die Straße teilweise mit Schnee teilweise feucht ohne Schnee und in zwei weiteren Videos



wechselte der Untergrund auch von trockener zu feuchter Straße. Die Lichtverhältnisse änderten sich in keinem Video vor dem Training. Die meisten Ereignisse finden bei trockener Straße und Luft am Tage statt. Das zeigt, dass die Auswahl der Videos gut gelungen ist, da das so auch den allgemeinen Straßenverkehr widerspiegelt. Relativ hoch ist jedoch auch der Anteil feuchter und nasser Straßen während der potenziell kritischen Ereignisse. In wenigen Fahrten lag Schnee während der Einsatzfahrt. Eis oder starke Verschmutzungen auf der Straße sind in der Stichprobe nicht aufgetreten. Hinsichtlich der Luftverhältnisse fanden alle Fahrten nur bei trockenem Wetter oder leichtem Regen statt. In Dunkelheit sind verhältnismäßig mehr Vorfahrtsregelungen aufgetaucht, dafür wurden in der Dämmerung und eingeschränktem Tageslicht mehr Überholvorgänge ausgeführt. Abbildung 53 verdeutlicht die Verteilung der Witterungsverhältnisse während der potenziell kritischen Situationen.

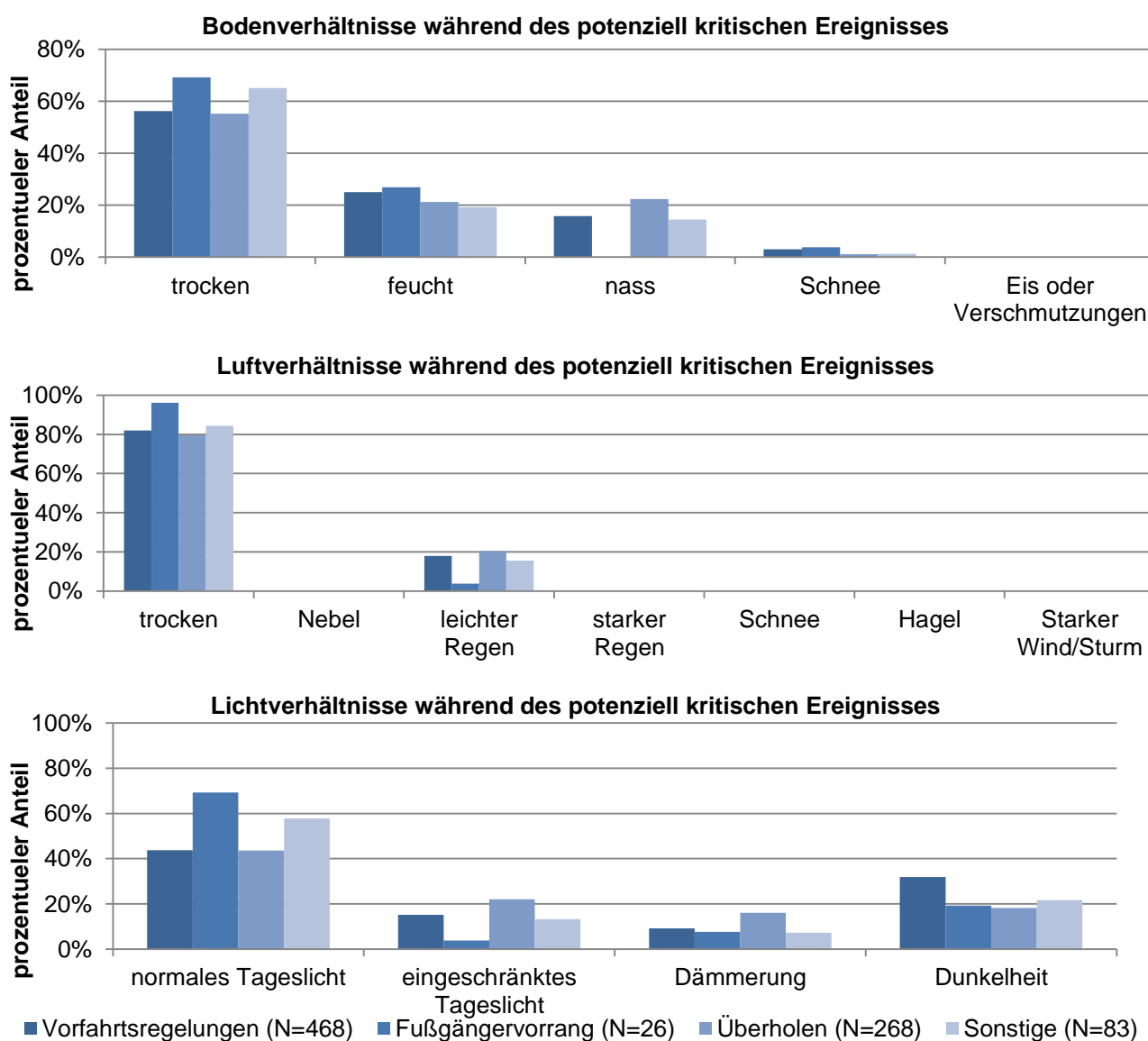


Abbildung 53: Witterungs- und Sichtverhältnisse während der potenziell kritischen Ereignisse in Abhängigkeit der Ereigniskategorie

### 7.3.3 Kategorien potenziell kritischer Situationen

Im Folgenden werden die potenziell kritischen Ereignisse hinsichtlich der *Vorfahrtsregelungen* betrachtet (vgl. Abbildung 54). Der Großteil der Ereignisse findet an Kreuzungen ohne Beschilderung (154 Fälle) und roten Ampeln statt (149 Fälle), gefolgt von Vorfahrtsregelungen mit Beschilderung (99 Fälle). Danach

kommt das Fahren in einen und im Kreisverkehr (40 Fälle). Die gelbe Ampel kommt mit 9 Fällen sogar seltener vor als Situationen, in denen der Einsatzfahrer eigentlich Vorfahrt hat, jedoch aufgrund der Verkehrsdichte oder anderer Situationen zu einer Fahrtanpassung gezwungen wird oder besonders aufpassen muss (17 Fälle).

In den meisten Fällen sind direkt vor der Kreuzung keine oder nur wenige Fahrzeuge, so dass die Fahrtroute selten blockiert ist. Für die Verkehrsdichte zählte jeweils der Fahrstreifen mit den wenigsten Fahrzeugen, wenn das Einsatzfahrzeug diesen Weg nutzte. So kann es beispielsweise sein, dass zwei Fahrstreifen voll mit Fahrzeugen sind, eine jedoch durch die übrigen Verkehrsteilnehmer frei gelassen wurde, da sie das Einsatzfahrzeug rechtzeitig wahrgenommen haben. In dem Fall gilt der Weg als frei. In nur fünf Fällen gab es explizite Probleme vor Kreuzungen, an den Fahrzeugen vorbei zu kommen. Wenn ohne Probleme in den Gegenverkehr ausgewichen werden kann, zählt die Situation auch als Möglichkeit, voran zu kommen. Dann wurde jedoch auch der Gegenverkehr kodiert (vgl. Abbildung 54).

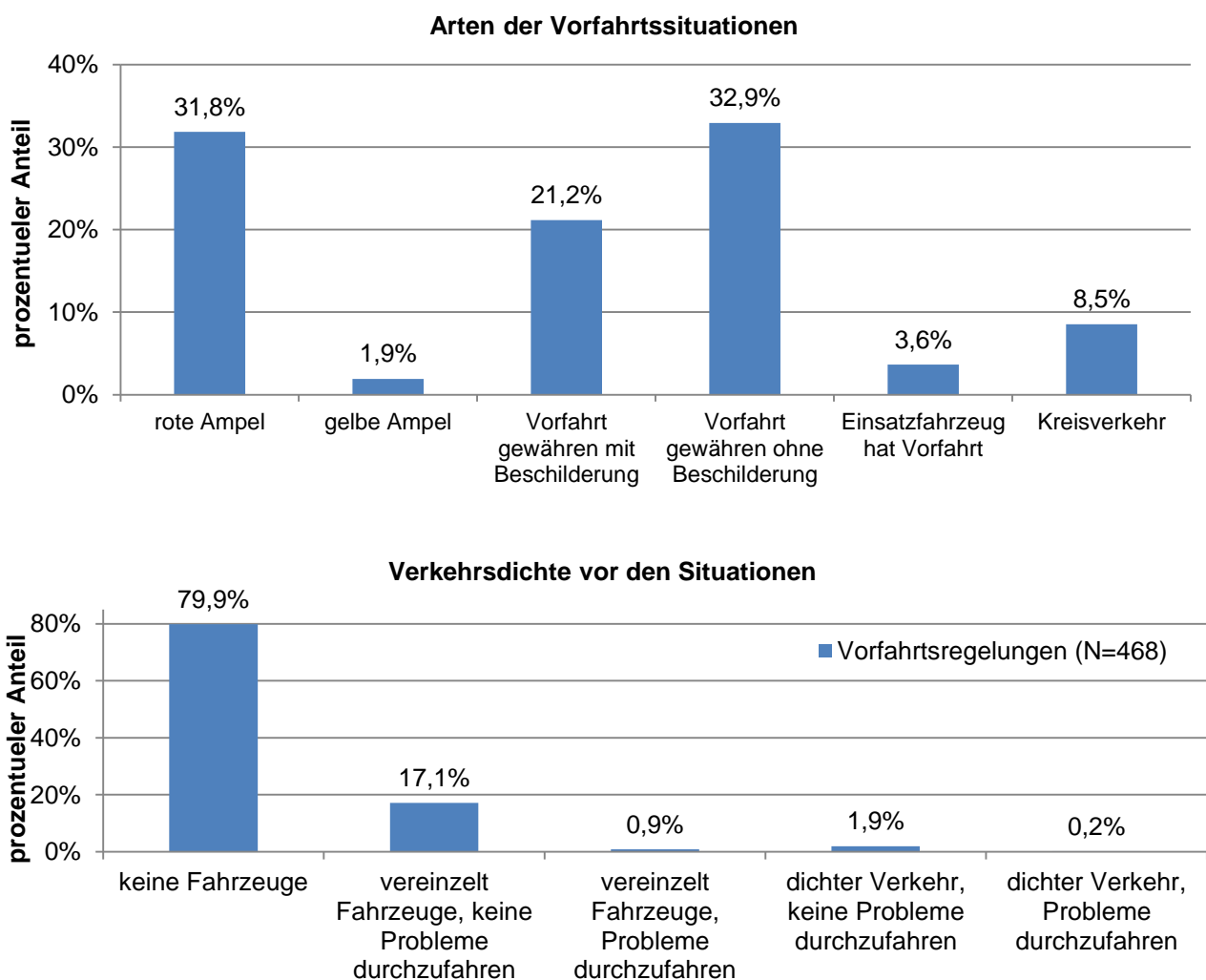


Abbildung 54: Vorfahrt: Arten und Verkehrsdichte vor den potenziell kritischen Situationen

An Kreuzungen interessant ist auch der Querverkehr, zu dem sowohl Fahrzeuge von rechts und links zählen, als auch Fußgänger und Radfahrer, die kreuzen oder über die Straße wollen, in die abgebogen wird. Aber auch der Gegenverkehr kann relevant sein, wenn beispielsweise in diesen ausgewichen wird oder nach links abgebogen werden muss. Diese beiden Variablen werden nur kodiert, wenn sie relevant sind, also beispielsweise links abgebogen oder gerade über eine Kreuzung gefahren wird, daher sind die Zahlen

insgesamt geringer. Dennoch ist auch hier zu sehen, dass in den meisten Situationen keine anderen Verkehrsteilnehmer vorhanden sind und wenn doch, diese oft bereits stehen. Immerhin in 14,3% der Situationen mit, in denen der Gegenverkehr relevant sein könnte, hält dieser an, wenn der Einsatzfahrer dessen Spur braucht. Nur in knapp über 5% der Fälle fahren oder laufen andere Verkehrsteilnehmer weiter und behindern so entweder das Einsatzfahrzeug oder lenken es ab, indem auf diese wiederum gesondert geachtet wird (vgl. Abbildung 55).

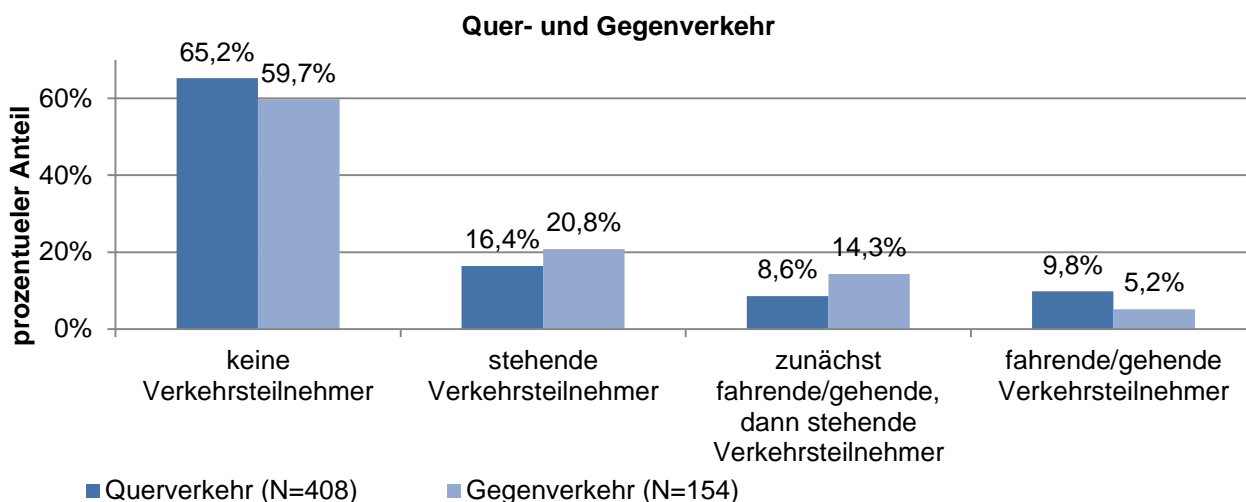


Abbildung 55: Vorfahrtssituationen: Ausprägungen des Quer- und Gegenverkehrs

Nachfolgend wird noch beschrieben, wie die Einsatzfahrer auf die potenziell kritischen Situationen reagieren (vgl. Tabelle 66). Insgesamt 145-mal (30,9%) reagieren die Einsatzfahrer gar nicht und fahren wie vor dem Ereignis weiter. 12 dieser Ereignisse sind rote Ampeln, die meisten Situationen sind beschilderte und unbeschilderte Kreuzungen (insgesamt 109). Die Bewertung der Reaktion ist jedoch unabhängig der gefahrenen Geschwindigkeit. Wenn beispielsweise schon vor der Kreuzungssituation sehr langsam gefahren wird, wird „keine Reaktion“ genauso kodiert, wie wenn jemand mit hoher Geschwindigkeit über eine Kreuzung fährt und nicht abbremst. Insgesamt am häufigsten wird aufgrund der Vorfahrtssituationen gebremst (268-mal bzw. 57,3%), wobei hier die Stärke des Bremsvorgangs nicht beurteilt wird. Bis zum Stillstand wurde nur sechsmal (1,3%) abgebremst. Ebenfalls sechsmal sind die Einsatzfahrer ausgewichen (1,3%) und 42-mal haben sie gebremst und sind ausgewichen (9,0%). Nur einmal wurde während einer Vorfahrtssituation gewendet (0,2%). Tabelle 66 gibt einen Überblick, welche Reaktion auf welche Art der Vorfahrtssituation wie häufig vorgekommen ist.

Tabelle 66: Häufigkeiten der Reaktionen auf die Arten der Vorfahrtssituationen

	Rote Ampel	Gelbe Ampel	Vorfahrt gewähren mit Beschilderung	Vorfahrt gewähren ohne Beschilderung	Einsatzfahrzeug hat Vorfahrt	Kreisverkehr
keine Reaktion	8,1%	55,6%	35,4%	48,1%	23,5%	37,5%
Ausweichen	2,0%	11,1%	0,0%	1,3%	0,0%	0,0%
Bremsen	69,1%	33,3%	58,6%	47,4%	52,9%	55,0%
Anhalten	0,0%	0,0%	3,0%	1,9%	0,0%	0,0%
Ausweichen & Bremsen	20,8%	0,0%	3,0%	0,6%	23,5%	7,5%
Bremsen & wenden	0,0%	0,0%	0,0%	0,6%	0,0%	0,0%
Summe der Fälle	149	9	99	154	17	40

Lediglich 26 Vorfahrtregelungen an *Fußgängerampeln* (zweimal rot, einmal grün) oder *Fußgängerüberwegen* (23-mal) wurden vor dem Training gefunden. Die Verkehrsdichte vor den Ampeln/Überwegen war nur zweimal mit vereinzelt Fahrzeugen, jeweils einmal mit bzw. ohne Probleme durchzufahren, sonst gab es keine Fahrzeuge. Insgesamt 20-mal gab es dabei keine Fußgänger, weitere sechsmal haben die vereinzelt Fußgänger dem Einsatzfahrzeug Platz gemacht. Acht Situationen, in denen der Gegenverkehr relevant hätte sein können, verliefen ohne Gegenverkehr, in den anderen spielte er keine Rolle. Die Einsatzfahrer sind 17-mal ohne Veränderung ihrer Fahrweise über Fußgängerüberwege gefahren. Achtmal wurde gebremst und einmal gebremst und ausgewichen.

268 *Überholvorgänge* wurden hinsichtlich potenziell kritischer Situationen kodiert. Dabei wurden alle Vorgänge, in denen zumindest teilweise auf die Gegenfahrbahn gefahren wird, aufgenommen. Die Art des Überholens wurde eingeteilt in Kurven oder Geraden, die entweder überschaubar oder unüberschaubar waren. Zusätzlich gab es gesondert die Kategorie Überholen in einer Stausituation, die vor dem Training in der gezogenen Stichprobe nur einmal auftrat (0.4%). Die häufigsten Überholvorgänge wurden in überschaubaren Situationen vollzogen (insgesamt 88.1%), aber auch immerhin 7.1% der Situationen waren unüberschaubare Kurven und 4.5% unüberschaubare Geraden.

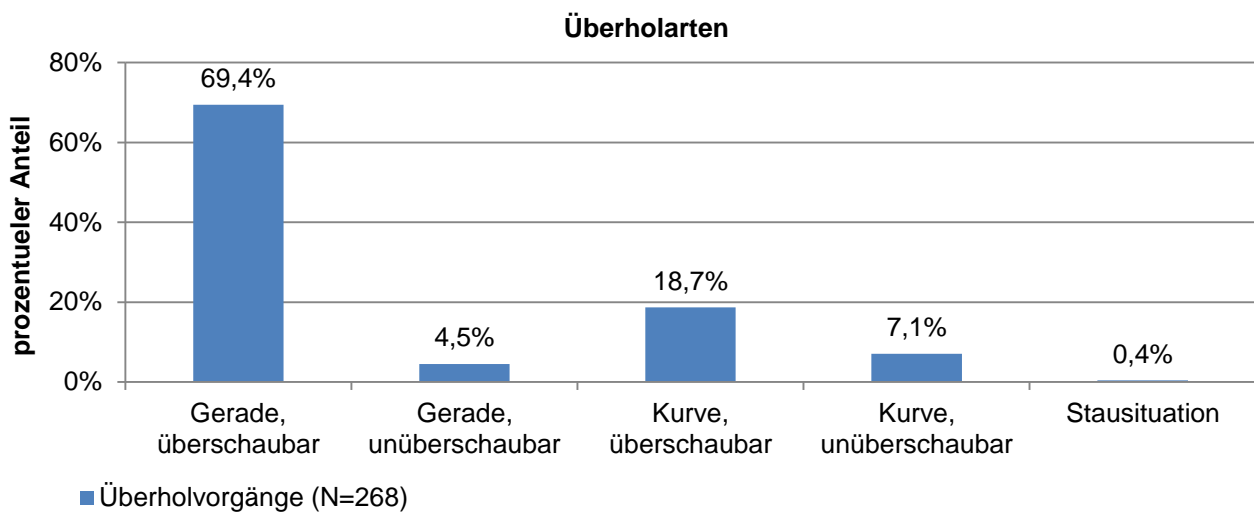


Abbildung 56: Arten von Überholvorgängen

In den meisten Fällen wurden nur wenige Fahrzeuge (eins bis drei) überholt (87.7%), gefolgt von vier bis neun Fahrzeugen (11.6%). Nur zweimal musste eine längere Kolonne von 10 oder mehr Fahrzeugen überholt werden (0.7%). Häufiger sind die zu überholenden Fahrzeuge selbst gefahren (63.4%), dabei ist jedoch nicht kodiert, wie schnell die zu überholenden Fahrzeuge waren oder ob diese durch möglichst weit rechts fahren oder blinken angedeutet haben, dass die das Einsatzfahrzeug wahrgenommen haben. In 36.6% der Überholvorgänge standen die Fahrzeuge, die überholt wurden. Wenn mehrere Fahrzeuge überholt wurden, zählte der Vorgang als fahrende Fahrzeuge, wenn mindestens eines der Fahrzeuge fuhr. Beim Überholen ist neben dem Fahrtrichtungsverkehr besonders der Gegenverkehr relevant um zwischen potenziell kritischen und kritischen Ereignissen unterscheiden zu können bzw. die Höhe der zusätzlichen Aufmerksamkeit bestimmen zu können. Abbildung 57 verdeutlicht die Kodierung des Gegenverkehrs. In den meisten Situationen war kein Gegenverkehr vorhanden, als überholt wurde (43.7%), direkt darauf folgend gab es jedoch fahrenden Gegenverkehr (27.2%) oder stehenden Gegenverkehr (18.7%). Nur sehr selten wurde rechts (1.5%) oder zwischen den Spuren (2.2%) überholt. In 6.7% der Überholungen gab es eine bauliche Trennung zum Gegenverkehr, so zum Beispiel immer auf Autobahnen.

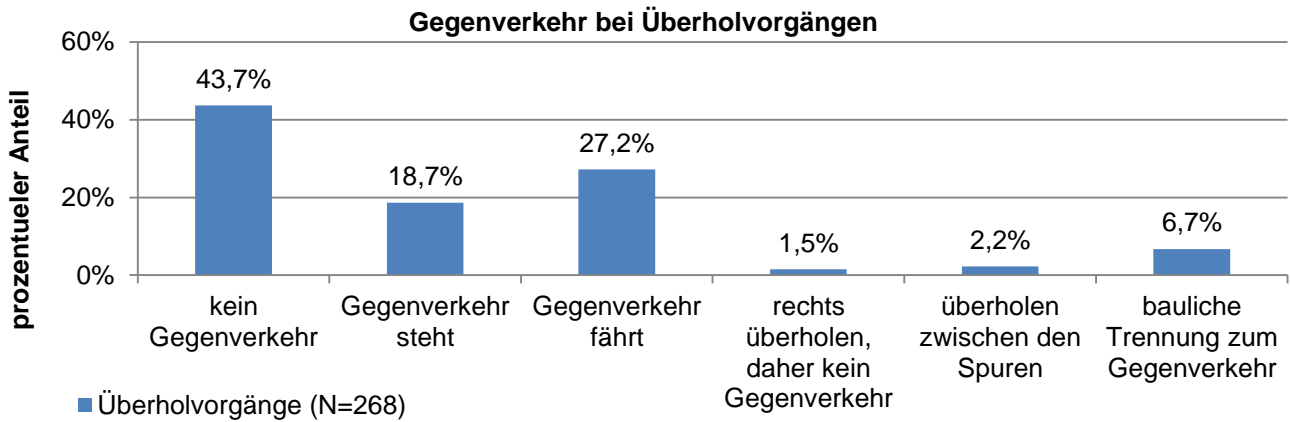


Abbildung 57: Kodierung des Gegenverkehrs bei Überholvorgängen

Die Reaktionen auf das Ereignis Überholen waren am häufigsten Ausweichen und Bremsen (45.9%) sowie Ausweichen (36.9%), häufig, um zum Überholen anzusetzen. In 11.9% der Überholmanöver wurde gebremst. Nur sehr selten musste komplett angehalten werden (0.4% Anhalten, 0.7% Ausweichen und Anhalten). In 4.1% der Situationen erfolgte keine Reaktion, weil das Einsatzfahrzeug langsam in den Gegenverkehr fuhr oder die übrigen Verkehrsteilnehmer so weit nach rechts gefahren sind, dass kein Ausweichen nötig war. Die folgende Tabelle gibt einen Überblick, welche Reaktion bei welchen Überholvorgängen wie häufig vorgekommen ist:

Tabelle 67: Häufigkeiten der Reaktionen auf die Arten der Überholsituationen

	Gerade, überschaubar	Gerade, unüberschaubar	Kurve, überschaubar	Kurve, unüberschaubar	Stausituation
keine Reaktion	3,2%	8,3%	6,0%	5,3%	0,0%
Ausweichen	43,0%	25,0%	24,0%	21,1%	0,0%
Bremsen	9,7%	25,0%	16,0%	15,8%	0,0%
Anhalten	0,0%	8,3%	0,0%	0,0%	0,0%
Ausweichen & Bremsen	43,5%	33,3%	52,0%	57,9%	100,0%
Ausweichen & Anhalten	0,5%	0,0%	2,0%	0,0%	0,0%
Summe der Fälle	186	12	50	19	1

Fiel eine potenziell kritische Situation nicht in die Vorfahrtsregelung (für Kreuzungen oder Fußgänger) oder in Überholvorgänge, wurde sie als *sonstige Kategorie* eingestuft. In den Vorsichtungen wurden bereits einige Arten sonstiger Situationen festgelegt, zum Beispiel das Entgegenkommen eines anderen Einsatzfahrzeuges mit Sondersignal oder Tiere, die vor das Fahrzeug gelaufen/geflogen sind. Diese beiden Arten tauchten jedoch beispielweise in den zufällig gezogenen Videos vor dem Training nicht mehr auf. Die Fahrt gegen die vorgeschriebene Fahrtrichtung kam am häufigsten vor (26.5%), gefolgt von der Behinderungen durch andere Verkehrsteilnehmer (24.1%). Zu durchfahrende Engstellen kamen mit 15.7% genauso oft vor, wie sich zu verfahren bzw. zu weit zu fahren. Die letztgenannte Situation ist für den Straßenverkehr im ersten Moment keine kritische Situation, kann jedoch beim Einsatzfahrer erhöhte Beanspruchung auslösen und so im Nachhinein aufgrund von beispielweise Wendemanövern oder der geteilten Aufmerksamkeit aufgrund der Suche des Einsatzortes zu potenziell kritischen Situationen führen. 4.8% der sonstigen potenziell kritischen Situationen fallen auf Spurwechsel. Diese sind entweder Wechsel auf Bus-/Bahnspuren, wenn die Ein- oder Ausfahrt aus diesen mit besonderer Vorsicht durchgeführt werden

müssen (beispielsweise, wenn dabei eine andere Fahrbahn überfahren werden muss) oder z.B. Wechsel von der Straße auf Grasstreifen oder Fußgängerwege. Unter den elf sonstigen Situationen fallen zum Beispiel eine sich vor dem Einsatzfahrzeug schließende Ausfahrts-Schranke, Kinder, die über die Straße laufen oder die Alarmierung während der Fahrt, die eine Fahrtänderung erzwang.

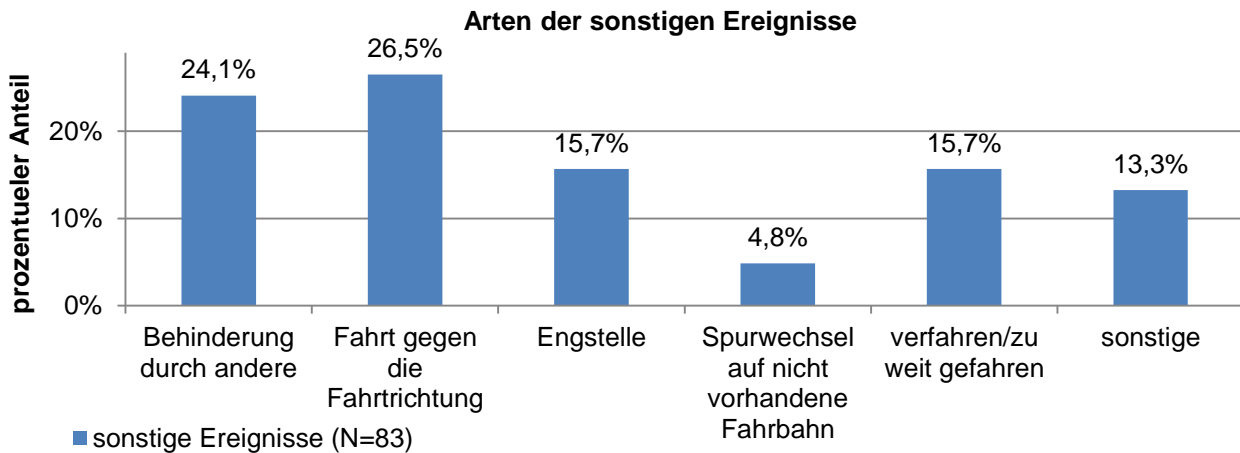


Abbildung 58: Arten von sonstigen potenziell kritischen Ereignissen

Hinsichtlich der Verkehrsdichte gab es in den meisten Fällen der sonstigen potenziell kritischen Ereignisse keine Fahrzeuge in Fahrtrichtung (69.9%). In weiteren 13.3% bzw. 1.2% gab es vereinzelte oder viele Fahrzeuge, die allerdings keine Probleme machten, durchzufahren. In insgesamt 15.6% der Fälle behinderten die Fahrzeuge in Fahrtrichtung das Einsatzfahrzeug (12.0% vereinzelte Fahrzeuge, 3.6% dichter Verkehr). Der Quer- und Gegenverkehr wurde nur kodiert, wenn er für das Ereignis Relevanz haben konnte. Insgesamt 25 sonstige potenziell kritische Situationen wurden hinsichtlich des Querverkehrs kodiert, wobei in den meisten Fällen keine Verkehrsteilnehmer vorhanden waren (60%). In jedoch immerhin 16% der kodierten Querverkehre gab es übrige Verkehrsteilnehmer, die weiter gefahren oder gelaufen sind und das Einsatzfahrzeug nicht wahrnahmen. In weiteren 20% wurde das Einsatzfahrzeug erst spät wahrgenommen. Bei der Kodierung des Gegenverkehrs (insgesamt 44 Fälle) kamen ebenfalls häufig keine Verkehrsteilnehmer (36.4%) vor. Stehende Verkehrsteilnehmer (25%) und zunächst fahrende, dann stehende Verkehrsteilnehmer (27.3%) kamen etwa gleichhäufig vor. In 11.4% der Fälle hat der Gegenverkehr das Einsatzfahrzeug nicht wahrgenommen oder nicht reagiert (vgl. Abbildung 59).

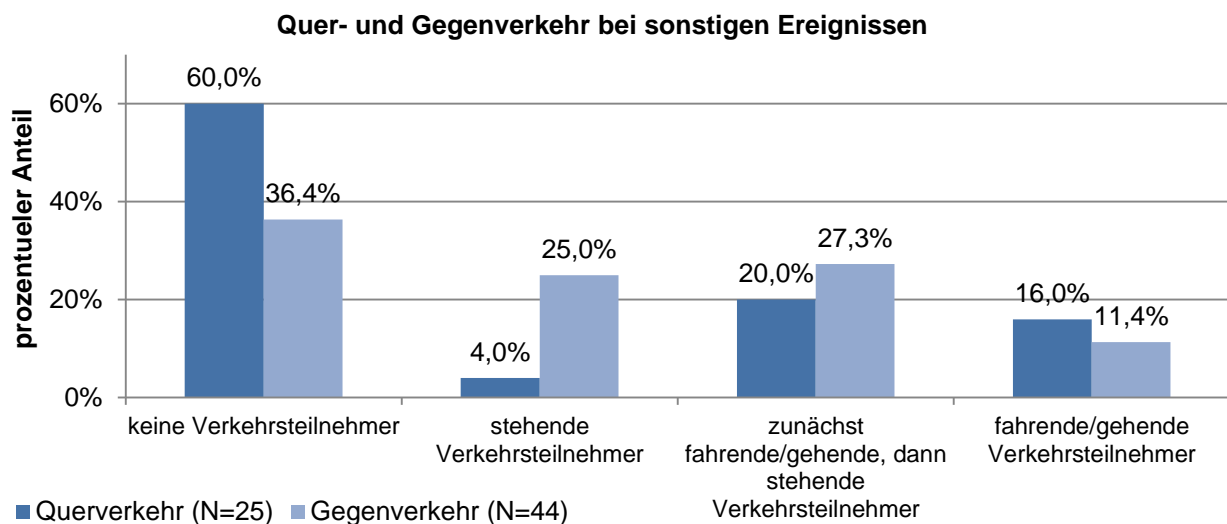


Abbildung 59: Ausprägung des Quer- und Gegenverkehrs bei den sonstigen potenziell kritischen Ereignissen

Die Reaktionen auf die sonstigen Ereignisse sind wie auch die Art der Ereignisse stärker verteilt als bei den anderen drei Kategorien (vgl. Tabelle 68). Am häufigsten kommen bremsen und ausweichen sowie Kombinationen aus diesen Reaktionen vor (insgesamt 71.1%). In 14.5% der Fälle wird die Fahrtroute geändert/gewendet, in 13.3% erfolgt gar keine Reaktion. Den Beschleunigungsvorgang abubrechen findet nur einmal statt (1.2%).

Tabelle 68: Häufigkeiten der Reaktionen auf die Arten der sonstigen Situationen

	Behinderung durch andere Verkehrsteilnehmer	Fahrt gegen die Fahrtrichtung	Engstelle	Spurwechsel auf nicht vorhandene Fahrbahn	verfahren/zu weit gefahren	sonstige
keine Reaktion	0,0%	31,8%	15,4%	25,0%	0,0%	9,1%
Ausweichen	5,0%	0,0%	7,7%	25,0%	0,0%	0,0%
Bremsen	35,0%	31,8%	46,2%	0,0%	7,7%	36,4%
Anhalten	25,0%	18,2%	0,0%	0,0%	23,1%	27,3%
Beschleunigung unterbrechen	5,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Wenden/Route ändern	0,0%	0,0%	7,7%	0,0%	61,5%	27,3%
Ausweichen & Bremsen	30,0%	0,0%	7,7%	50,0%	0,0%	0,0%
Ausweichen & Anhalten	0,0%	4,5%	15,4%	0,0%	0,0%	0,0%
Bremsen & anhalten	0,0%	13,6%	0,0%	0,0%	0,0%	0,0%
Bremsen & wenden	0,0%	0,0%	0,0%	0,0%	7,7%	0,0%
Summe der Fälle	20	22	13	4	13	11

## 7.4 Einflussfaktoren auf Beanspruchung im Rettungsdienst

Wie bereits in einigen Abschnitten berichtet wurde, hängen die Beanspruchungen, die auch in der Resultateebene ausgewertet wurden, häufig mit weiteren Faktoren zusammen. So können beispielsweise gesundheitsbezogene Daten, wie der Raucherstatus oder die sportliche Aktivität, Erfahrung oder einsatzbezogene Variablen sich auf die erlebte Beanspruchung auswirken. In diesem Abschnitt werden Zusammenhänge zwischen Kontrollvariablen und Beanspruchungsvariablen dargestellt.

Die Auswertungen erfolgten jeweils regressionsanalytisch. Mittels einer linearen Regression wurde der Einfluss der demographischen, gesundheits-, erfahrungs- und schichtbezogenen Daten ermittelt. Diese vier Bereiche wurden nacheinander der Berechnung hinzugefügt. Die demographischen Daten enthalten das Alter und Geschlecht, welches wie bisher kontrolliert wurde.

Die *gesundheitsbezogenen Daten* enthalten den Raucherstatus (1=noch nie geraucht, 2=nicht mehr, 3=gelegentlich, 4=regelmäßiges Rauchen), die sportliche Aktivität (1=nie bis 5=täglich), den BMI, Herz-Kreislaufkrankungen (0=nein, 1=ja) und die Anzahl der regelmäßig einzunehmenden Medikamente (ohne Kontrazeptiva). *Erfahrungsvariablen* sind die gleichen, die auch in Abschnitt 3.4.5 bereits zur Berechnung genutzt wurden (Qualifikation, Rettungsdienst Erfahrung in Monaten, Fahrzeiten im Rettungsdienst und Führerscheinbesitz in Jahren). Im letzten Schritt wurde überprüft, welche *schichtbezogenen Daten* eine Rolle spielen könnten. Überprüft werden die Startzeit der Schicht bzw. des EKG, die Schichtlänge, die Art der Schicht und die Anzahl der Einsätze während der Schicht. Die demographischen Daten werden mittels Einschluss im ersten Block eingefügt die anderen Variablen werden mittels schrittweiser Selektionsverfahren hinzugefügt. So können alle Variablen, die einen signifikanten Einfluss (Irrtumswahrscheinlichkeit von 5%) auf die Beanspruchung haben, herausgefunden werden. Hinsichtlich der EKG-Erhebungen wurden die mittlere Herzfrequenz, die Anzahl der Tachykardien während der gemessenen Schicht sowie der Befund

ausgewertet. Der Befund ist dreigeteilt zwischen „unauffälliges EKG“, erhöhte mittlere Herzfrequenz und der Empfehlung einen Arzt aufzusuchen.

Tabelle 69: Einflussfaktoren auf die EKG-Messungen zum Zeitpunkt T1

Abhängige Variable	Mittlere Herzfrequenz (N=91)			Anzahl der Tachykardien (N=90)			Befund des EKG (N=91)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.026	.026		.013	.013		.036	.036
...Alter	-0.04			-0.07			-0.10		
...Geschlecht	-0.14			-0.07			-0.13		
Schritt 2: Gesundheit		.081 <sup>+</sup>	.056 <sup>*</sup>					.105 <sup>*</sup>	.069 <sup>*</sup>
...Raucherstatus	0.24 <sup>*</sup>								
...körperliche Aktivität							-0.26 <sup>*</sup>		
Schritt 3: Erfahrung		.140 <sup>*</sup>	.059 <sup>*</sup>						
...Qualifikation	0.25 <sup>*</sup>								
Schritt 4: Schichtdaten									

Berechnung: Schrittweise Regression unter Einschluss des Schritt 1 zum Zeitpunkten T1, dargestellt sind jeweils nur die Variablen, die einen signifikanten Einfluss auf die abhängige Variable haben; Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Die Schichtdaten, das Alter und Geschlecht haben auf keine der drei abhängigen Variablen einen Einfluss. Die mittlere Herzfrequenz wird durch den Raucherstatus beeinflusst und unterscheidet sich zwischen den Qualifikationen. Je eher die Probanden rauchen, desto höher ist die mittlere Herzfrequenz. Hinsichtlich der Qualifikation zeigen höher ausgebildete Rettungsdienstmitarbeiter höhere Herzfrequenzen. Die Anzahl der Tachykardien wird durch keine der gemessenen Kontrollvariablen beeinflusst. Der Befund hängt lediglich mit der körperlichen Aktivität zusammen. Probanden, die häufiger Sport trieben, erhielten seltener die Empfehlung, einen Arzt aufzusuchen.

Um die Kontrollvariablen auch für die Beanspruchungsmaße der Schicht berechnen zu können und um die Verzerrung durch Mittelwerte oder ungepaarte Stichproben zu vermeiden, wurde für den Messzeitpunkt T1 die zweite Schicht genutzt, um wie bei den EKGs Einflussfaktoren bestimmen zu können. Die zweite Schicht wurde aus dem Grund genutzt, um evtl. Veränderungen der Beanspruchung aufgrund der zum Zeitpunkt T1 neu eingeführten Messung während der Schicht zu umgehen und dennoch eine ausreichend große Gruppe an Probanden zu haben. Eine Übersicht der Ergebnisse befindet sich in Tabellen im Anhang B.



## 8. Zusammenfassung der Ergebnisse

Im folgenden Abschnitt werden die Ergebnisse der Trainingsevaluation vor dem Hintergrund der aufgestellten Hypothesen zusammengefasst. Dabei werden neben den Ergebnissen des kontrollierten Hauptdesigns jeweils auch die Ergebnisse zur Replizierbarkeit und Nachhaltigkeit dargestellt. Weitere, nicht hypothesenprüfende Ergebnisse werden im Anschluss dargestellt.

### **Hypothese 1 zur Reaktionsebene: Die Teilnehmer nehmen das Training als sinnvoll und hilfreich wahr.**

Alle der sechs getesteten Variablen werden nach dem Training mit sehr gut bis gut bewertet. Diese Effekte finden sich sowohl in den Interventionsgruppen wie auch in der Wartegruppe, sind also replizierbar. Die Trainingsteilnehmer nehmen das Training hinsichtlich der Inhalte, dem Trainer, der Rahmenbedingungen, positiv und hinsichtlich der Anwendbarkeit als sinnvoll und hilfreich wahr. Sie würden es weiterempfehlen und geben Gesamtnoten zwischen 1.5 und 2. In der längerfristigen Betrachtung wird das Training in allen drei getesteten Variablen Anwendbarkeit, erfolgte Weiterempfehlung und der Gesamtnote weiterhin durchschnittlich als gut wahrgenommen, d.h. die Bewertung nimmt etwas ab und stabilisiert sich auf diesem Niveau. Damit ist die Reaktion auf das Training auch nachhaltig positiv, das Training wird sowohl kurzfristig als auch längerfristig als sinnvoll und hilfreich wahrgenommen. Die erste Hypothese kann vollständig angenommen werden.

### **Hypothese 2a zur Lernebene: Die Teilnehmer lernen durch das Training relevantes Wissen zur Verkehrssicherheit.**

Das Wissen wurde mithilfe einer Gesamtvariablen, einem Wissenstest mit zwei Parallelversionen, getestet. Die Trainingsteilnehmer wiesen einen signifikanten Wissenszuwachs mit einer mittleren Effektstärke im Vergleich zu Nichtteilnehmern auf. Über die Replizierbarkeit und Nachhaltigkeit kann keine Aussage gemacht werden, da nur zu den Zeitpunkten T1 und T2 Messungen durchgeführt wurden. Der Wissenszuwachs bleibt auch unter Hinzunahme der Dropouts stabil. Hypothese 2a kann damit vollständig angenommen werden, die Teilnehmer lernen durch das Training relevantes Wissen zur Verkehrssicherheit.

### **Hypothese 2b zur Lernebene: Die Teilnehmer verändern durch das Training ihre Einstellung hin zu einer höheren Bedeutung von Verkehrssicherheit, höherem Verantwortungsbewusstsein und geringerer Risikobereitschaft.**

Zur Testung der Hypothese 2b (Einstellungsänderungen durch das Training) wurden insgesamt 15 Aspekte getestet, nur in einer dieser 15 fanden sich signifikante Veränderungen: Die Risikowahrnehmung bei Normalfahrten erhöhte sich in den Interventionsgruppen. Die Veränderung weist eine geringe Effektstärke auf, konnte aber in der Wartegruppe mit hoher Effektstärke repliziert werden. Allerdings sind die Effekte nicht nachhaltig, die Erhöhung der Risikowahrnehmung nimmt langfristig wieder ab und die Unterschiede zwischen letztem und erstem Messzeitpunkt sind nicht mehr signifikant, wenn auch deskriptiv am Ende immer noch höher als zu Beginn, d.h. ein Trend in die gewünschte Richtung bleibt. Unter Hinzunahme von Dropouts zeigen die Mittelwerte der Risikowahrnehmung bei Normalfahrten den gleichen Verlauf und lassen auf belastbare Ergebnisse schließen. Insgesamt kann die Hypothese 2b damit nur für einen speziellen Bereich der Einstellungsänderung angenommen werden.

### **Hypothese 3 zur Verhaltensebene: Das Training hat einen positiven Effekt auf das reale Fahrverhalten der Teilnehmer.**

Das Fahrverhalten wurde insgesamt mithilfe von 26 Variablen, die zum Teil voneinander abhängig sind bzw. auf die gleichen Ausgangswerte zurückgreifen, gemessen. Zwei dieser Variablen, die mittlere und maximale Geschwindigkeit, wurden durch das Training signifikant verringert. Die Verringerung wies eine geringe Effektstärke auf und konnte durch die Analysen der Wartegruppe nicht repliziert werden, obwohl die Mittelwerte auch in dieser Gruppe nach dem Training etwas geringer waren als vor dem Training.

Langfristig konnte ebenfalls eine Verringerung der mittleren und maximalen Geschwindigkeit festgestellt werden, somit wirkt sich das Training nachhaltig positiv auf die gefahrene Geschwindigkeit aus. Im Trend bleiben die Effekte auch bei den unverbundenen Analysen unter Hinzunahme der Dropouts bestehen, so dass von belastbaren Werten ausgegangen werden kann. Allerdings verweisen die Werte hier auch auf potenzielle Spezifika der ausgeschiedenen Teilnehmer. Insgesamt kann die Hypothese 3 nur für die mittlere und maximale Geschwindigkeit bei Sondersignalfahrten angenommen werden, nicht aber für alle anderen Parameter des realen Fahrverhaltens.

**Hypothese 4a zur Resultatebene: Das Training vermindert die erlebte Beanspruchung der Teilnehmer während der Einsatzfahrten.**

Insgesamt 26 psychische und physische, übergreifende und schichtbezogene Variablen wurden hinsichtlich der Hypothese 4a erhoben. Keine der erhobenen Variablen zeigte eine Veränderung durch das Training, weder in den Interventionsgruppen noch hinsichtlich der Replizierbarkeit in den Analysen der Wartegruppe. Auch die Vermutung, dass die Beanspruchung erst längerfristig sinken könnte, kann nicht gezeigt werden – in den Vergleichen zwischen erstem und letztem Messzeitpunkt finden sich keine Unterschiede. Dies ändert sich auch nicht bei Hinzunahme der Dropouts. Hypothese 4a muss damit verworfen werden, die Probanden erleben durch das Training keine Verminderung der erlebten Beanspruchung.

**Hypothese 4b zur Resultatebene: Das Training hat einen positiven Effekt auf typische Kennwerte einer Einsatzfahrt.**

Hinsichtlich der typischen Kennwerte konnte lediglich die Dauer der Einsatzfahrten als Näherungsvariable („Proxy“) für die „Hilfsfrist“ belastbar erfasst werden. Diese ändert sich nicht durch das Training, weder im Hauptdesign, noch für die Wartegruppe oder auf längere Sicht. Auch die Hinzunahme ändert daran nichts. Die Hypothese 4b kann demzufolge ebenfalls nicht angenommen werden. Jedoch ist es insgesamt als positiv zu bewerten, dass sich die Dauer der Einsatzfahrten durch das Training trotz verringerter Geschwindigkeiten nicht verlängert hat. Das Training hat demzufolge weder einen positiven noch einen negativen Effekt auf die Dauer der Einsatzfahrten.

**Hypothese 5: Erfahrung hat einen Einfluss auf den Trainingseffekt.**

Fünf Erfahrungsvariablen wurden in die Moderationsberechnungen für die Trainingseffekte aufgenommen, zwei betreffen die Erfahrung im Rettungsdienst und drei die Fahrerfahrung. Kurzfristig beeinflusst nur die Qualifikation (Rettungsdienstenerfahrung) den Wissenszuwachs und die maximale gefahrene Geschwindigkeit. Höher Qualifizierte profitieren kurzfristig etwas mehr vom Training als niedriger qualifizierte Mitarbeiter. Langfristig zeigen Mitarbeiter mit längerer Erfahrung im Rettungsdienst eine etwas höhere Abnahme der maximalen gefahrenen Geschwindigkeit. Erfahrung im Straßenverkehr zeigt in keiner der Analysen einen systematischen Einfluss auf den Trainingseffekt. Die Hypothese 5 kann damit insgesamt nur in Teilen angenommen werden.

Zusammengefasst konnten damit auf jeder Ebene - mit Ausnahme der Resultatebene – hypothesenkonforme Effekte des Trainings in einzelnen Variablen gefunden werden. Daneben fanden sich unabhängig vom Evaluationsdesign einige weitere interessante Befunde, die im Folgenden ebenfalls kurz zusammengefasst werden.

Bei vielen erfassten Variablen gab es einen Alters- und/oder Geschlechtseinfluss. Zehn der gemessenen 15 Variablen der Lernebene zeigen Unterschiede zwischen den älteren und jüngeren Teilnehmern. Die jüngeren zeigten beispielsweise geringere Risikowahrnehmungen und risikoreichere verkehrsrelevante Persönlichkeitseigenschaften, aber auch ein höheres verkehrsrelevantes Wissen. Hinsichtlich der Fahrweise konnten häufiger (7 der 18 Fahrprofildaten) Unterschiede des Geschlechts gefunden werden. Männer fuhren insgesamt mit höheren Beschleunigungen als Frauen. In den gemessenen Resultaten gab es

ebenfalls Alters- und Geschlechtseinflüsse, jedoch deutlich seltener und unspezifischer als in der Lern- und Verhaltensebene.

Auch Einflussfaktoren auf den individuellen Trainingserfolg wurden betrachtet. Individuell und gruppenbezogen scheint der Spaß am Fahren mit Sondersignal eine der relevantesten Variablen zu sein. Personen, die nicht gern mit Sondersignal fahren, zeigen individuell kurzfristig höhere Trainingserfolge. Die Trainingsgruppe profitiert – sowohl kurz- als auch langfristig –, wenn wenigstens eine Person dabei ist, die nicht gern mit Sondersignal fährt. Individuell scheinen daneben das Alter, längerer Führerscheinbesitz, und wenige Jahreskilometer kurzfristig zu einem Trainingserfolg zu führen. Für den längerfristigen Erfolg konnten auf individueller Ebene Einflussfaktoren gefunden werden. Hinsichtlich der organisationsbezogenen Variablen zeigen sich kurzfristig eher kontraintuitive Einflussfaktoren: weniger Arbeitsplatzautonomie, geringere interne Kooperation, geringere soziale Unterstützung, höherer patientenbezogener Stress, höherer Zeitdruck und strafender Umgang mit Fehlern führt eher zu Trainingserfolg. Langfristig wirken sich jedoch eher der weniger strafende Umgang mit Fehlern und geringerer Zeitdruck positiv auf den Trainingserfolg aus. Die Gruppenzusammensetzung sollte neben dem genannten Faktor eher heterogen sein und wenigstens eine Person mit weniger Freude am Fahren mit Sondersignal sollte in der Trainingsgruppe dabei sein.

Für die erlebte Beanspruchung während der Schichten im Rettungsdienst finden sich kaum Einflüsse der Erfahrung oder Gesundheit der Probanden. Am häufigsten wirken sich die Art der Schicht, der Schichtbeginn und die Anzahl der Einsätze über die Schicht auf die erlebte Beanspruchung aus. Je später die Schicht beginnt und je mehr Einsätze abgearbeitet werden müssen, desto höher werden die Beanspruchungen erlebt.

Blaulichtfahrten zum Einsatzort weisen insgesamt trotz gleicher mittlerer Geschwindigkeit höhere Beschleunigungen und maximale Geschwindigkeiten auf als Blaulichtfahrten mit einem Patienten im Fahrzeug. Den größten Einfluss auf die Fahrweise während der Schichten haben die Schichtdauer und die Anzahl der Einsätze sowie die geistigen Anforderungen über die Schichten. Entgegen der ursprünglichen Vermutung konnten kaum Einflüsse des Einsatzstichwortes auf die Fahrweise gefunden werden. Auch die rettungsdienst- oder verkehrsbezogene Erfahrung zeigt so gut wie keinen Einfluss auf die Fahrweise. Für Einsatzfahrten konnte zudem gezeigt werden, dass die Häufigkeit des Auftretens potenziell kritischer Situationen geringer zu sein scheint als bisher angenommen. In der Studie der BAST kamen kritische Situationen etwa alle 19 Sekunden vor (Unterkofler & Schmiedel, 1994). In der vorliegenden Stichprobe kommt es etwa alle 37 Sekunden zu einem potenziell kritischen Ereignis, also einem Ereignis, das eine erhöhte Aufmerksamkeit fordert. Die häufigsten dieser Ereignisse betreffen Vorfahrtregelungen (ca. 55%) gefolgt von Überholvorgängen (ca. 32%). Die meisten Situationen treten innerorts, auf einspurigen Straßen, während trockener Witterung und bei Tageslicht auf.

## 9. Diskussion und praktische Implikationen

Bevor im Folgenden die praktischen Implikationen der vorliegenden Studie diskutiert werden, werden zunächst einige methodische Aspekte genauer erläutert. Das Ziel der Studie, durch ein Training für Blaulichtfahrer eine höhere Verkehrssicherheit zu erzielen, wurde mit einem komplexen Studiendesign und aufwändigen objektiven und subjektiven Messmethoden auf vier verschiedenen Evaluationsebenen untersucht. Ein kontrolliertes Wartegruppen-Design lässt zu, dass Effekte tatsächlich auf das Training zurückgeführt werden können, gesonderte Messungen der Wartegruppe erlauben die Testung der Replizierbarkeit der Befunde und eine Nacherhebung nach einem längeren Zeitraum hilft, die Nachhaltigkeit von Effekten zu testen. Darüber hinaus konnten durch Tests mit unverbundenen Stichproben Hinweise zur weiteren Belastbarkeit der Ergebnisse und zu potenziellen Verzerrungen der Stichprobe gewonnen werden.

Letzteres stellt einen ersten zu diskutierenden methodischen Aspekt dar – die Frage der Stichprobenverzerrung in der Studie. In der Akquise (vgl. Abschnitt 3.4.1) wurden alle Rettungsdienstorganisationen bestimmter geographischer Regionen angesprochen und konnten zumindest mit einzelnen Verbänden auch von der Teilnahme überzeugt werden, so dass hier vermutlich höchstens von unsystematischen Problemen ausgegangen werden kann. Problematischer ist hingegen die Situation bei der – entsprechend dem Ethikvotum notwendigen – freiwilligen Teilnahme der Rettungsdienstmitarbeiter. Hier ist zunächst generell von Verzerrungen auszugehen, da einer nicht zufällig gezogenen Stichprobe immer das potenzielle Problem der Unausgewogenheit inhärent ist. Dieser sogenannte Sampling Bias beschreibt den Fakt, dass nicht alle intendierte Teilnehmer einer Studie o.ä. die gleiche Wahrscheinlichkeit haben, für die Teilnahme ausgewählt zu werden oder teilzunehmen (vgl. z.B. Deeks et al., 2003; Grimes & Schulz, 2002; Tripepi et al. 2008, 2010). Bei Interventionen, wie dem hier untersuchten Training, besteht dieser Wahrscheinlichkeitsunterschied vor allem darin, dass es zur Selbstselektion kommt - vermutlich waren nur Personen, die ein gewisses Interesse an der Thematik haben, Verkehrssicherheit bei Sondersignalfahrten als wichtig ansehen und dafür auch vergleichsweise hohe Messaufwände in Kauf zu nehmen, bereit, sich für die Teilnahme zur Verfügung zu stellen, so dass in der vorliegenden Studie von höher motivierten Teilnehmern auszugehen ist als in der allgemeinen Gruppe von Rettungsdienstmitarbeitern. Zusätzlich zu dieser ersten Einschränkung bei der vorliegenden Stichprobe findet sich möglicherweise auch – wie zum Teil an hohen Dropout-Raten und vorsichtig zu interpretierenden Hinweisen aus Analysen mit unverbundenen Stichproben – ein potenziell systematischer Teilnehmerschwund im Sinne einer sogenannten Sample Attrition (z.B. Jüni & Egger, 2005): Dieses Problem betrifft den Punkt, dass das Aussteigen aus einer Studie oder das nicht (vollständige) Ausfüllen von Fragebogen oder ähnlichem, nicht zufällig geschieht, sondern besonders diejenigen Teilnehmer betrifft, für die Intervention nicht passt oder funktioniert. Im Fall der vorliegenden Studie scheint es zumindest Hinweise darauf zu geben, dass vor allem Rettungsdienstmitarbeiter im Laufe des Projekts ausgestiegen sind, die eher rasantere Fahrer sind. Beide Aspekte der Stichprobenverzerrung, sollten sie zutreffend sein, würden zunächst einmal bedeuten, dass die vorgestellten Befunde nur für interessierte und möglicherweise bereits relativ ruhige Fahrer gelten. Aussagen über die Wirksamkeit des Trainings für Personen, die weniger an Verkehrssicherheit interessiert sind, können damit vermutlich nicht oder nur bedingt getroffen werden.

Ähnlich der Stichprobenverzerrung bei den teilnehmenden Personen, muss für die aufwändigen Schichtmessungen eine Verzerrung der erfassten Schichten vermutet werden: Wie in Abschnitt 3.2.4 dargelegt, wurden Schichten zwar im Prinzip zufällig nach Dienstplänen und opportunistisch nach Geräteverfügbarkeiten und ähnlichem gemessen und sollten vergleichsweise repräsentativ sein, andererseits sind gerade besonders „stressige“ Schichten mit vielen Einsatzfahrten bereits zu Beginn manchmal nicht oder nur mit Minimalangaben gemessen worden, da die Teilnehmer den Aufwand der Erhebung in diesen Schichten nicht erbringen konnten oder wollten. Damit findet sich auch auf der Seite der Erhebungen ein potenziell systematischer Fehler, der dazu führen könnte, dass die aufgefundenen

Effekte eher für „normal“ belastende Sondersignalfahrten gelten. Einflüsse auf die Unterschiede zwischen den Gruppen und den Messzeitpunkte sind aber nicht zu erwarten, da dieser Effekt gleichmäßig für alle Messungen zutreffen müsste. Da zudem viele Schichten und Fahrten pro Teilnehmer erhoben wurden und eine Erfassung verschiedener weiterer Einflussfaktoren vorliegt (vgl. Abschnitt 3.2.1.5), kann insgesamt davon ausgegangen werden, dass dieser Effekt eher gering ist.

Neben den genannten Stichprobenverzerrungen, die in Evaluationsstudien von Trainings und Interventionen nahezu nicht zu verhindern sind, findet sich in der vorliegenden Studie ein weiteres methodisches Problem, das diskutiert werden muss – das Problem multiplen Testens. Um die Effekte des Trainings möglichst breit auf den verschiedenen Ebenen potenziellen Lernens zu testen, wurden viele unterschiedliche Konstrukte betrachtet und an immer derselben Stichprobe statistisch betrachtet. Dieses Vorgehen führt zu der sogenannten Alphafehler-Kumulierung oder  $\alpha$ -Fehler-Inflation – dem Phänomen, dass durch multiples Testen in derselben Stichprobe die Wahrscheinlichkeit erhöht wird, signifikante Befunde zu finden, die tatsächlich  $\alpha$ -Fehler sind (d.h. Zurückweisung der Nullhypothese, obwohl sie in Wirklichkeit wahr ist). In den vorliegenden Analysen wurde dieses Problem noch dadurch verschärft, dass keine der üblichen Alphaadjustierungsmethoden (z.B. Benjamini & Hochberg, 1995; Holm, 1979) angewendet wurde, sondern sogar eine Irrtumswahrscheinlichkeit von 10% angelegt wurde, um den  $\beta$ -Fehler zu verkleinern, d.h. um keine real existierende Effekte zu übersehen. Auch wenn dieser Aspekt beachtet werden sollte, spricht eine Reihe von Punkten dafür, dass die aufgefundenen Effekte nicht statistische Zufallsartefakte sind, sondern reale Unterschiede abbilden. Das erste und oft auch in der Literatur benannte Argument, um diesem Problem zu entgehen, ist hypothesengeleitetes Testen (vgl. z.B. Victor, Elsässer, Hommel & Blettner, 2010), das in der vorliegenden Studie stattgefunden hat. Das zweite Argument ist die Art der aufgefundenen Effekte. Zum einen treten alle Befunde in der erwarteten Richtung auf und zum anderen sind es nahezu immer die gleichen Variablen über die verschiedenen Analyseschritte (Hauptdesign, Wartegruppenanalyse, Langfristanalysen), in denen sich signifikante Effekte finden, so dass nahezu ausgeschlossen werden kann, dass es sich bei den Ergebnissen um Zufallsbefunde handelt.

Vor dem Hintergrund dieser methodenkritischen Anmerkungen bleibt die Frage bestehen: Was bringt das Training? Und lohnt es sich, das Training mehr „in die Fläche“ zu bringen? Kurz: Welche praktischen Implikationen haben die Projektergebnisse für alle Beteiligten und besonders für die gesetzlichen Unfallversicherer? Betrachtet man die Zusammenfassung der Befunde in Abschnitt 8, dann könnte der erste Impuls sein, eher von einem Misserfolg oder einer Nichtwirksamkeit des Trainings auszugehen, bei genauerer Betrachtung scheint dies aber nicht haltbar zu sein: Die – für alle praktischen Zwecke – zentralste Ebene des Verhaltens der Rettungsdienstfahrer, zeigt durchgängig Verringerungen in den mittleren und maximalen Fahrgeschwindigkeiten als Effekte des Trainings. Die unmittelbaren Verringerungen der mittleren Durchschnittsgeschwindigkeit von knapp 55 km/h auf 53.5 km/h und der durchschnittlichen Maximalgeschwindigkeit von knapp 118 km/h auf knapp 113 km/h sowie die langfristig gefundenen Verringerungen um knapp 3 km/h in der mittleren Geschwindigkeit und 4.5 km/h in den maximalen Geschwindigkeiten erscheinen dabei zunächst nicht viel, aber schon bei trockener Fahrbahn und mit den Faustformeln für PKW berechnet verringern sich Brems- und Anhaltewege um mehrere Meter, betrachtet man etwa die längerfristige Senkung der Maximalgeschwindigkeit, kommen die Fahrer nun bei einer Normalbremsung etwa 12 Meter früher zum Stehen, bei einer Gefahrenbremsung sind es immerhin noch 6.5 Meter, und während sie jetzt schon stehen, würden sie mit der Maximalgeschwindigkeit vor dem Training noch 8.2 km/h schnell sein. Ob sich diese Werte unmittelbar in geringere Unfallzahlen übersetzen lassen, lässt sich im Rahmen des Projekts nicht klären – da bei den Unfallursachen allgemein „nicht angepasste Geschwindigkeiten“ immer in den obersten Rängen liegen und ursächlich für die meisten Toten im Straßenverkehr sind (Statistisches Bundesamt, 2016), wird aber die Relevanz einer Geschwindigkeitsreduktion für die Verkehrssicherheit deutlich. In diesem Zusammenhang zeigen die Ergebnisse einen weiteren erfreulichen Befund – trotz der niedrigeren Geschwindigkeiten verlängern sich die Einsatzfahrten nicht signifikant, d.h. die Retter sind genauso schnell dort, wo sie gebraucht werden, wie vor dem Training.

Weniger gut quantifizierbar sind die Effekte eines erhöhten Risikobewusstseins und einer verbesserten Wissensgrundlage, dennoch sollten auch diese hilfreich für ein angemessenes Verhalten im Straßenverkehr sein.

Verschiebt man den Fokus von Mittelwerten und statistischen Unterschieden hin zu individuellen Entwicklungen und Veränderungen durch das Training, bleibt auch hier die Bewertung der Wirksamkeit und Nützlichkeit des Trainings letztlich eine normative Frage. Längst nicht jeder Teilnehmer hat vollständig, auf verschiedenen Lernebenen oder in einzelnen Teilaspekten von dem Training profitiert und nicht für alle Personen bleiben positive Entwicklungen über die Zeit bestehen. Dieser Befund ist kein Alleinstellungsmerkmal des untersuchten Trainings und findet sich, wenn in Studien überhaupt betrachtet, in den unterschiedlichsten Bereichen von Fort- und Weiterbildungen. In diesem Zusammenhang stellt sich auch schnell die Frage nach der Dosis-Wirkungs-Beziehung (vgl. Meyers & Hessler, 2007; Wittchen & Jacobi, 2007), die erst in neuerer Zeit auch für psychologisch-medizinische Untersuchungen eine Rolle spielt. Das Training „Verkehrssicherheit bei Einsatzfahrten“ ist ein eintägiger Kurs, in dem neben der Wissensvermittlung zweimal von Teilnehmern selbst im Simulator gefahren wird und die Fahrten der anderen Teilnehmer gemeinsam besprochen werden. Es erscheint nicht nur nicht verwunderlich, dass diese sehr geringe Dosis eher eine kleine Anzahl von Effekten produziert, sondern auch, dass längst nicht jeder von dieser Intervention profitiert. So erscheint es durchaus passend, dass bei den individuellen Voraussetzungen erfolgreicher Teilnahme zum einen Personen zu finden sind, die noch nicht viele Fortbildungen in diesem Bereich gemacht haben und im Hinblick auf psychologische Lernkurven (vgl. z.B. Ritter & Schooler, 2002) damit einen steileren Lernanstieg zu erwarten haben als Personen, die bereits einiges gelernt haben. Und, zum anderen, dass eher höherqualifizierte Personen profitieren, von denen angenommen werden kann, dass die individuellen Lernvoraussetzungen, wie kognitive Fähigkeiten, stärker ausgeprägt sind. Eine häufigere Wiederholung oder Auffrischung des Trainings könnte sicherlich in letzterer Gruppe zu mehr Effekten führen, allerdings sollte im Sinne der Ersteren und der „klassischen“ Dosis-Wirkungsbeziehungen auch beachtet werden, dass ein „zuviel“ ein Problem werden kann.

Betrachtet man die Ergebnisse zum individuellen Erfolg und die Rückmeldungen zu den Trainings, lassen sich auch einige Hinweise und Empfehlungen zur Organisation der Trainings ableiten. Ein lediglich technisches, allerdings häufiger genanntes und möglicherweise für einen Teil der Unterschiede zwischen den Erhebungsorten zuständiges Thema, sind die Simulatoren, die nicht mehr als zeitgemäß empfunden wurden und sicherlich einen kleinen Einfluss darauf haben, in welchem Maß sich die Teilnehmer auf das Training einlassen. Deutlich wichtiger scheint hingegen die Zusammensetzung der Trainingsgruppen zu sein - es scheint (und dies hat im Hinblick auf das angestrebte Voneinander-Lernen großen Sinn) hilfreich zu sein, die Trainings möglichst heterogen zusammenzusetzen und darauf zu achten, dass auch Personen in den Gruppen sind, denen das Fahren mit Sondersignal keinen großen Spaß macht – sozialpsychologisch kann so ein Minoritäteneinfluss, eine Minderheitenmeinung, in einer Gruppe dazu führen, dass intrapsychische kognitive Konflikte der Gruppenmitglieder zu echten Veränderungen führen (vgl. Konversionstheorie; Moscovici & Lage, 1969, 1976). Teilnehmer mit geringem Spaß an Sondersignalfahrten sollten daher leichter „begeisterte“ Fahrer dazu bringen können, Gefahren und Risiken ernst zu nehmen.

Fasst man alle Ergebnisse des Projekts zusammen, muss abschließend konstatiert werden, dass das Training belastbare und positive, wenn auch kleine, Effekte erzielt und ggf. durch einige Maßnahmen noch in seiner Wirksamkeit verbessert werden kann. In einem so wichtigen Gebiet wie der Verkehrssicherheit bei Einsatzfahrten, in dem neben dem Fahrer nicht nur andere Verkehrsteilnehmer, sondern auch Hilfsbedürftige von einer sicheren Fahrweise abhängen, sollten auch solche Effekte als Erfolg verbucht werden.

## 10. Kommunikation und Umsetzung der Ergebnisse

Für die Verbreitung der Ergebnisse in Wissenschaft und Praxis ist eine Reihe von Aktivitäten geplant oder bereits auf den Weg gebracht. Bereits *während* des Projektes wurde darauf geachtet, dass Ergebnisse entsprechend weitergegeben wurden. So wurde jeder teilnehmenden Rettungsdienstorganisation angeboten, die Ergebnisse der eigenen Mitarbeiter präsentiert zu bekommen, wenn genügend Personen aus dem entsprechenden Dienst teilgenommen hatten, um den Datenschutz garantieren zu können. Waren die Teilnehmerzahlen zu klein, wurde eine Rückmeldung mit aus mehreren Organisationen zusammengefassten Ergebnissen angeboten und durchgeführt, wenn die betroffenen Rettungsdienste mit einer Zusammenlegung einverstanden waren. Wurde auch dies nicht gewünscht, konnte alternativ eine Präsentation der Gesamtergebnisse zum jeweiligen Zeitpunkt erfolgen. Von den 26 teilnehmenden Ortsverbänden wurden bei einem die Gesamtergebnisse präsentiert, bei 13 Verbänden zusammengelegte Ergebnisse und bei neun Verbänden die organisationsspezifischen Ergebnisse. Da innerhalb des Projekts auch eine erhebliche Anzahl von gesundheitsbezogenen Aspekten erfasst wurden, erhielten die Teilnehmer auf Wunsch nach Abschluss der letzten Erhebungen der jeweiligen Erhebungswelle auch eine individuelle zeh- bis zwölfseitige Gesundheitsrückmeldung zu ihren längsschnittlichen Befunden in den Bereichen Wohlbefinden im Arbeitszusammenhang – Irritation und Burnout, Allgemeines Wohlbefinden – depressive Verstimmung, körperliche Beschwerden, selbsteingeschätzte Arbeitsfähigkeit und kardiovaskuläre Befunde der Schicht-EKGs mit einer Einordnung der Ergebnisse und Empfehlungen für den Umgang mit den eigenen Werten (wie z.B. eine fachärztliche Abklärung). Insgesamt lehnten nur zwei Teilnehmer diese Rückmeldung explizit ab, 168 Personen erhielten eine solche individuelle Befundung.

Auch auf wissenschaftlicher Seite wurden bereits während der Projektlaufzeit einzelne Aspekte in der Fachöffentlichkeit zur Diskussion gestellt: So wurde erstmalig auf dem Doktorandenworkshop der Fachgruppe Verkehrspsychologie der Deutschen Gesellschaft für Psychologie 2013 das Studiendesign diskutiert und auf den Wissenschaftlichen Jahrestagungen der Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V. in den Jahren 2014, 2015 und 2016, das Studiendesign, die Erfassung von verkehrssicherheitsrelevantem Wissen und die emotionale Erschöpfung, Depersonalisation und Depression von Rettungsdienstmitarbeitern präsentiert. In 2015 wurde auf der Deutschen Fachtagung Notfallpsychologie die Frage, wie sich Verkehrsunfälle bei Rettungsdienstmitarbeitern langfristig auf Wohlbefinden, Irritation und Burnout auswirken, diskutiert. In 2017 wurden schließlich auf der Conference of Experimental Psychologists (TeaP) die Abhängigkeit der Ermüdung von Rettungsdienstmitarbeitern von den Einsätzen pro Schicht und auf der Tagung der Fachgruppe Arbeits-, Organisations- und Wirtschaftspsychologie der Deutschen Gesellschaft für Psychologie die wechselseitige Abhängigkeit der Reaktions- und Lernebenen bei der Trainingsevaluation vorgestellt.

In der *Zukunft* werden diese wissenschaftlichen Aktivitäten im Hinblick auf die Gesamtergebnisse des Projekts erweitert und fortgeführt. So werden die Wirksamkeitsanalysen erstmalig im März 2018 auf der 58. Wissenschaftlichen Jahrestagung der Deutschen Gesellschaft für Arbeitsmedizin und Umweltmedizin e.V. vorgestellt und sollen in Folge auch auf weiteren thematisch passenden Tagungen und Kongressen präsentiert werden (z.B. Workshop Psychologie der Arbeitssicherheit und Gesundheit, PASIG, Gemeinsames Symposium der Deutschen Gesellschaft für Verkehrsmedizin e.V. und der Deutschen Gesellschaft für Verkehrspsychologie e.V.). Darüber hinaus soll die Kommunikation in die Wissenschaft vor allem über nationale und internationale peer-review Publikationen erfolgen. Geeignete Journals zur Veröffentlichung der gesamten Trainingsevaluation sind u.a. Accident Analysis and Prevention, Transportation Research Part F: Traffic Psychology and Behaviour, oder Journal of Transportation Safety & Security. Für einzelne Aspekte der Methodenentwicklung, Fahrprofil- und Videoauswertungen kommen z.B. die Zeitschrift für Verkehrssicherheit, Advances in Health Sciences Education oder das Journal of Continuing Education in the Health Professions in Frage.

Auch die Kommunikation der Ergebnisse in die Praxis soll auf verschiedenen Wegen geschehen. Zum Ersten werden alle teilnehmenden Rettungsdienstorganisationen diesen Abschlussbericht erhalten sowie eine Kurzzusammenfassung der Befunde, die sie auf Wunsch über eigene Kanäle (z.B. Intranet, Newsletter) weiterverbreiten können. Des Weiteren wurde bereits und wird auch weiterhin den Dachorganisationen der beteiligten Verbände angeboten, in breiterem Rahmen (z.B. bei Sitzungen und Tagungen der (Arbeitsgemeinschaft der) Zweckverbände oder der Leiter Rettungsdienst) das Training und die Evaluationsergebnisse zu präsentieren. Dies soll auch in Richtung der gesetzlichen Unfallversicherer, etwa über das Sachgebiet "Feuerwehren und Hilfeleistungsorganisationen" der Deutschen Gesetzlichen Unfallversicherung stattfinden. Um die Praxis in der Fläche anzusprechen ist zudem angedacht, Ergebnisse der Studie komprimiert in Zeitschriften zu publizieren, die eine hohe Verbreitung im Rettungsdienst haben, wie z.B. „Rettungsdienst: Zeitschrift für präklinische Notfallmedizin“ oder „Im Einsatz: Zeitschrift für Einsatzkräfte im Katastrophenschutz“. Gemeinsam mit den Kooperationspartnern und Simulatorinhabern Deutscher Verkehrssicherheitsrat e.V. und Institut für Arbeit und Gesundheit der Deutschen gesetzlichen Unfallversicherung e.V. ist eine Nutzbarmachung der Auswertungen der Fahrprofil- und Videodaten für die Erarbeitung neuer/weiterer Szenarien für die Trainingssimulatoren bereits in Planung. Inwieweit der Deutsche Verkehrssicherheitsrat die Projektergebnisse etwa im Rahmen einer Broschüre zusätzlich öffentlich verbreiten möchte, ist noch zu diskutieren.



## Literaturverzeichnis

- Åkerstedt, T. & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52, 29-37.
- Albertsson, P. & Sundström, A. (2011). Evaluation of insight training of ambulance drivers in Sweden using dart, a new e-learning tool. *Traffic Injury Prevention*, 12, 621-629.
- Baron, R. M. & Kenny D. A. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- BAuA (2016). *Arbeitszeitreport Deutschland 2016*. Dortmund: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Bech, P. (2004). Measuring the dimensions of psychological general well-being by the WHO-5. *Quality of Life Newsletter*, 32, 15-16.
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 57 (1), 125–133.
- Bockting, S. (2007). *Verkehrsunfallanalyse bei der Nutzung von Sonder- und Wegerechten gemäß StVO*. Hamburg: Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege.
- Brähler, E., Hinz, A. & Scheer, J.W. (2008). *GBB-24 Gießener Beschwerdebogen*. Göttingen: Hogrefe.
- Büssing, A. & Glaser, J. (2002). *Das Tätigkeits- und Arbeitsanalyseverfahren für das Krankenhaus – Selbstbeobachtungsversion (TAA-KH-S)*. Göttingen: Hogrefe.
- Büssing, A. & Perrar, K. M. (1992). Die Messung von Burnout. Untersuchung einer Deutschen Fassung des Maslach Burnout Inventory (MBI-D). *Diagnostica*, 38, 165-180.
- Cohen, P., Cohen, J., Aiken, L. S. & West, S. G. (1999). The Problem of Units and the Circumstance for POMP. *Multivariate Behavioral Research*, 34 (3), 315-346.
- Damm, L., Nachtergaele, C., Meskali, M. & Berthelon, C. (2011). The evaluation of traditional and early driver training with simulated accident scenarios. *Human Factors*, 53, 323-337.
- Deeks J. J., Dinnes J., D'Amico R., Sowden A. J., Sakarovitch C., Song F., Petticrew M. & Altman D. G. (2003). International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technology Assessment*, 7 (27), 1-173.
- Delhomme, P. (1991). Comparing one's driving with others': Assessment of abilities and frequency of offences. Evidence for a superior conformity of self-bias? *Accident Analysis & Prevention*, 23, 493-508.
- Deutscher Verkehrssicherheitsrat (2008). Mehr Sicherheit für Einsatzfahrer. *DVR Report*, 4/2008, 10.
- DIN EN ISO 10075-1 (2000). *Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung – Teil 1: Allgemeines und Begriffe*. Berlin: Beuth.
- Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. et al. (2006). *The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment*. (Report No. DOT HS 810 593). Washington, D.C.: National Highway Traffic Safety Administration.
- Dorn, L. & Barker, D. (2005). The effects of driver training on simulated driving performance. *Accident Analysis and Prevention*, 37, 63-69.
- Ericsson, K. A., Charness, N., Hoffman, R. R. & Feltovich, P. J. (2006). (Eds.). *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press.
- Eissfeldt, N. & Wagner, P. (2003). Effects of anticipatory driving in a traffic flow model. *The European Physical Journal B*, 33, 121-129.
- Gauger, J. F. (2011). *Eine kritische Betrachtung der Entwicklungen im deutschen Rettungswesen - Konzepte eines zukünftigen außerklinischen Versorgungssystems*. Karlsruhe: Karlsruher Institut für Technologie, Institut für Operations Research.
- Glaser, J. & Herbig, B. (2012). Modelle der psychischen Belastung und Beanspruchung. In DIN Deutsches Institut für Normierung e.V. (Hrsg.), *Psychische Belastung und Beanspruchung am Arbeitsplatz - Inklusive DIN EN ISO 10075-1 bis -3* (S. 17-27). Berlin: Beuth Verlag.
- Glaser, J., Seubert C., Hornung, S. & Herbig, B. (2015). The impact of learning demands, work-related resources, and job stressors on creative performance and health. *Journal of Personnel Psychology*, 14(1), 37-48.

- Goode, N., Salmon, P. M. & Lenné, M. G. (2013). Simulation-based driver and vehicle crew training: Applications, efficacy and future directions. *Applied Ergonomics*, 44, 435-444.
- Grasmick, H. G., Tittle, C. R., Bursik, R. J., Jr., & Arneklev, B. J. (1993). Testing the core empirical implications of Gottfredson and Hirschi's general theory of crime. *Journal of Research in Crime and Delinquency*, 30, 5-29.
- Grimes D. A. & Schulz K. F. (2002). Bias and causal associations in observational research. *Lancet*, 359, 248-252.
- Hacker, W. (2003). Action Regulation Theory: A practical tool for the design of modern work processes? *European Journal of Work and Organizational Psychology*, 12, 105-130.
- Hacker, W. (2005). *Allgemeine Arbeitspsychologie. Psychische Regulation von Wissens-, Denk- und körperlicher Arbeit*. Bern: Huber.
- Hacker, W. & Skell, W. (1993). *Lernen in der Arbeit*. Berlin, Bonn: Bundesinstitut für Berufsbildung.
- Hammer, A., Ernstmann, N., Ommen, O., Wirtz, M., Manser, T., Pfeiffer, Y. & Pfaff, H. (2011). Psychometric properties of the Hospital Survey on Patient Safety Culture for hospital management (HSOPS\_M). *BMC Health Services Research*, 11, 165-174.
- Hart, S. G. & Staveland, L. E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. S. Hancock & N. Meshkati (Hrsg.), *Human mental work-load* (pp. 139-183). New York: North-Holland.
- Henkel, V., Mergl, R., Kohnen, R., Maier, W., Möller, H. J. & Hegerl, U. (2003). Identifying depression in primary care: a comparison of different methods in a prospective cohort study. *British Medical Journal*, 326, 200-201.
- Herbig, B. & Glaser, J. (2013). *Kreativität und Gesundheit im Arbeitsprozess – Bestandsaufnahme, Intervention und Evaluation* (Abschlussbericht zum Projekt F 2244). Dortmund, Berlin, Dresden: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Herbig, B. & Glöckner, A. (2009). *Experts and Decision Making: First steps towards a unifying theory of decision making in novices, intermediates and experts*. (Preprints of the Max Planck Institute for Research on Collective Goods 2009/2). Bonn: Max Planck Institute for Research on Collective Goods.
- Herrnkind, K. (2017, 2. Juni). Alle paar Tage kracht es. Abgerufen von <https://www.stern.de/panorama/gesellschaft/unfaelle-mit-rettungswagen-2016-und-2017-7479332.html>
- Hessen: Jeden Tag 5 Blaulichtunfälle. (2015, 6. Oktober). Abgerufen von <http://www.fnp.de/rhein-main/blaulicht/Hessen-Jeden-Tag-5-Blaulicht-Unfaelle;art25945,1628437>
- Hobi, V. (1985). *Basler Befindlichkeitsskala*. Weinheim: Beltz Test Gesellschaft.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6 (2), 65–70.
- Horswill, M. S., Waylen, A. E. & Tofield, M. I. (2004). Drivers' Ratings of different components of their own driving skill: a greater illusion of superiority for skills that relate to accident involvement. *Journal of Applied Social Psychology*. 34 (1), 177-195.
- INM - Institut für Notfallmedizin und Medizinmanagement (2016). *Rettungsdienstbericht Bayern 2016*. München: Institut für Notfallmedizin und Medizinmanagement.
- Isler, R. B., Starkey, N. J. & Sheppard, P. (2011). Effects of higher-order driving skill training on young, inexperienced drivers' on-road driving performance. *Accident Analysis and Prevention*, 43, 1818-1827.
- Isler, R. B., Starkey, N. J., Sheppard, P. & Yu, C. (2007). Piloting a telemetric data tracking system to assess post-training real driving performance of young novice drivers. In L. Dorn (Ed.), *Driver Behaviour and Training* (pp. 17-29). London: Ashgate.
- Ivancic, K. & Hesketh, B. (2000). Learning from errors in a driving simulation: effects on driving skill and self-confidence. *Ergonomics*, 43, 1966-1984.
- Jarczok, M. N., Jarczok, M., Mauss, D., Koenig, J., Li, J., Herr, R. M. & Thayer, J. F. (2013). Autonomic nervous system activity and workplace stressors - A systematic review. *Neuroscience & Biobehavioral Reviews*, 37(8), 1810–1823
- Jost, C. (2012). *Sicher im Einsatz*. Unveröffentlichter Vortrag im Rahmen des Trainings „Verkehrssicherheit bei Einsatzfahrten“ am Institut für Arbeit und Gesundheit der DGUV, Dresden, August 2012.

- Jüni, P. & Egger, M. (2005). Empirical evidence of attrition bias in clinical trials. *International Journal of Epidemiology*, 34 (1), 87–88.
- Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology* 67, 332-339.
- Kirkpatrick, D. L. (1996). Great Ideas Revisited. *Training and Development Journal*, 50(1), 54-59.
- Kirkpatrick, D. L. (1998). *Evaluating training programs. The four levels* (2nd ed.). San Francisco: Berrett-Koehler Publishers.
- Kirkpatrick, J. D. & Kirkpatrick, W. K. (2016). *Kirkpatrick's Four Levels of Training Evaluation*. USA: ATD Press.
- Knoop, T. & Röthemeier, T. (2012, 18. Januar). Warum haben unsere Retter so oft Unfälle? Abgerufen von <http://www.bild.de/regional/hamburg/einsatzfahrzeuge/immer-mehr-verkehrsunfaelle-polizei-und-feuerwehr-22130068.bild.html>
- Krüger, H.-P. (2010). *Einsatzfahrten – Belastung und Stress der Einsatzkräfte*. Vortrag auf der Fachtagung „Sichere (Einsatz)-Fahrten bei Feuerwehr und Hilfeleistungsorganisationen“ am Institut für Arbeit und Gesundheit der DGUV, Dresden, Juni 2010.
- McKenna, F. P., Horswil, M. S. & Alexander, J. L. (2006). Does Anticipation Training Affect Drivers' Risk Taking? *Journal of Experimental Psychology: Applied*, 12, 1-10.
- Mohr, G., Müller, A. & Rigotti, T. (2005). Normwerte der Skala Irritation: Zwei Dimensionen psychischer Beanspruchung. *Diagnostica*, 51, 12-20.
- Mohr, G., Rigotti, T. & Müller, A. (2005). Irritation – Ein Instrument zur Erfassung psychischer Beanspruchung im Arbeitskontext. Skalen- und Itemparameter aus 15 Studien. *Zeitschrift für Arbeits- und Organisationspsychologie*, 49, 44-48.
- Molina, J. G., Sanmartín, J., Keskinen, E. & Sanders, N. (2007). Post-license education for novice drivers: Evaluation of a training programme implemented in Spain. *Journal of Safety Research*, 38, 357-366.
- Moscovici, S und Lage, E. (1976). Studies in social influence III: Majority versus minority influence in a group. *European Journal of Social Psychology*, 6, 149-174.
- Moscovici, S., Lage, E. & Naffrechoux, M. (1969). Influence of a consistent minority on the response of a majority in a color perception task. *Sociometry*, 32, 365-379.
- Müller, D. (2003). *Einsatzfahrten – ihre Risiken und Rechtsgrundlagen*. Vortrag auf dem 3. Internationalen Samariter-Forum vom 12.-14.09.2003. Radolfzell/Bodensee.
- Müller, D. (2010a). Aus- und Fortbildung von Einsatzfahrern im Rettungsdienst – ein Risiko für Notärzte? *Der Notarzt*, 26, 204-208.
- Müller, D. (2010b). *Einsatzfahrten – Checklisten zu Rechtmäßigkeit und Rechtsfolgen* (3. Aufl.). Stuttgart: Boorberg.
- Müller, D. (2011). Sichere Blaulichtfahrten: Die Bedeutung des Personals und der Ausbildung. *Rettungsdienst*, 34, 58-64.
- Müller, A., Petru, R. & Angerer, P. (2011). Cognitive Demands and the Relationship Between Age and Workload in Apron Control. *Aviation Space and Environmental Medicine*, 82, 26-33.
- Müller, A., Petru, R., Seitz, L., Englmann, I. & Angerer, P. (2010). The relation of cognitive load and pupillary unrest. *International Archives of Occupational and Environmental Health*, 84, 561-567.
- Myers, P. & Hessler, W. (2007, 30. April). Does 'the dose make the poison?' Extensive results challenge a core assumption in toxicology. *Environmental Health News*.
- Neukum, A., Lang, B. & Krueger, H.-P. (2003). A Simulator-Based Training for Emergency Vehicle Driving. In *Proceedings of the Driving Simulation Conference North America 2003*. USA: Dearborn, Michigan.
- Nübling, M., Stöbel, U., Hasselhorn, H.-M., Michaelis, M. & Hofmann, F. (2005). *Methoden zur Erfassung psychischer Belastungen – Erprobung eines Messinstruments (COPSOQ)*. Berlin: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Perrissol, S., Smeding, A., Laumond, F. & Le Floch, V. (2011). Effect of a road safety training program on drivers' comparative optimism. *Accident Analysis and Prevention*, 43, 478-482.
- Pieper-Nagel, M. (2011). *Evaluationsergebnisse zum simulationsgestützten Seminarkonzept „Verkehrssicherheit für Einsatzfahrer“*. Vortrag auf der Technology based Training for Drivers – Conference am Institut für Arbeit und Gesundheit der DGUV, Dresden, November 2011.
- Pieper-Nagel, M. & Wiegand, J. (2011). Nur wer ankommt, kann helfen. *DVR Report*, 2/2011, 13-15.

- Raheb, R. (2011). The Ripple Effect of a Traffic Collision. *Fire Engineering Magazine*, 164, 22-26.
- Ritter, F. E. & Schooler, L. J. (2002). *The learning curve*. In International Encyclopedia of the Social and Behavioral Sciences. Amsterdam: Pergamon.
- Robert-Koch-Institut (2010). *Daten und Fakten: Ergebnisse der Studie „Gesundheit in Deutschland aktuell 2009“* (Hrsg., Beiträge zur Gesundheits-berichterstattung des Bundes). Berlin: Robert-Koch-Institut.
- Rosenbloom, T., Shahar, A., Elharar, A. & Danino, O. (2008). Risk perception of driving as a function of advanced training aimed at recognizing and handling risks in demanding driving situations. *Accident Analysis and Prevention*, 40, 697-703.
- Schmiedel, R., Betzler, E. & Behrendt, H. (2004). *Bedarfsplanung im Rettungsdienst*. Berlin: Springer.
- Schmiedel, R. & Behrendt, H. (2011). *Leistungen des Rettungsdienstes 2012/13 – Analyse des Leistungsniveaus im Rettungsdienst für die Jahre 2012 und 2013*. Bundesanstalt für Straßenwesen (Hrsg.), Berichte der Bundesanstalt für Straßenwesen. Mensch und Sicherheit, Heft M 260. Bremen: Fachverlag NW.
- Schneider, C. B., Pilhatsch, M., Rifati, M., Jost, W. H., Wodarz, F., Ebersbach, G. et al. (2010). Utility of the WHO-Five Wellbeing Index as a Screening Tool for Depression in Parkinson's Disease. *Movement Disorders*, 25(6), 777-783
- Seipel, C. (2012). Deutsche Version der Self-Control Skala von Grasmick et. al. (1993). In A. Glöckner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 15.00. Bonn: GESIS.
- Spicher, B. & Hänsgen, K.-D. (2003). *TVP – Test zur Erfassung verkehrsrelevanter Persönlichkeitsmerkmale*. Bern: Verlag Hans Huber.
- Statistisches Bundesamt (2017). *Verkehr Verkehrsunfälle 2016* (Fachserie 8 Reihe 7). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2016). *Unfallentwicklung auf deutschen Strassen 2015*. Wiesbaden: Statistisches Bundesamt.
- Steinvoord, M. (2012). *GIS-gestützte Analyse von Fahrtgeschwindigkeiten unter Sonder- und Wegerecht* (Bachelorarbeit im Studiengang Gefahrenabwehr). Hamburg: Hochschule für Angewandte Wissenschaften.
- Tripepi G, Jager K. J., Dekker F. W., Wanner C. & Zoccali C. (2008). Bias in clinical research. *Kidney International*, 73(2), 148-53.
- Tripepi G., Jager K. J., Dekker F. W. & Zoccali C. (2010). Selection bias and information bias in clinical research. *Nephron Clinical Practice*, 115 (2), c94-9.
- Tronsmoen, T. (2010). Associations between driver training, determinants of risky driving behavior and crash involvement. *Safety Science*, 48, 35-45.
- Topp, C. W., Ostergaard S. D. & Sondergaard S. (2015) The WHO-5 Well-Being Index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84, 167-176.
- Unfallkasse Baden-Württemberg (2010). *Sicherheit auf Einsatzfahrten. Praxisnahe Übungen für Feuerwehr-Einsatzfahrer*. Stuttgart: Unfallkasse Baden-Württemberg.
- Ungerer, D. (2005). Anforderungen und Belastungen bei Einsatzfahrten. *Polizeispiegel*, 1/2, 40-42.
- Unterkofler, M. & Schmiedel, R. (1994). *Verbesserung der Sicherheit bei Sondersignaleinsätzen*. (Bericht zum Forschungsprojekt 8933 der Bundesanstalt für Straßenwesen Bereich Unfallforschung). Bergisch Gladbach: Bundesanstalt für Straßenwesen.
- Van der Hulst, M., Meijman, T. & Rothengatter, T. (1999). Anticipation and the adaptive control of safety margins in driving. *Ergonomics*, 42, 336-345.
- Varnskühler, B. (2011). *Simulatorgestütztes Training für Einsatzfahrer – Erste Erfahrungen bei der Unfallkasse des Bundes*. Vortrag auf dem internationalen Kongress für Arbeitsschutz und Arbeitsmedizin, Düsseldorf, Oktober 2011.
- Victor, A., Elsässer, A., Hommel, G. & Blettner, M. (2010). Wie bewertet man die p-Wert-Flut? *Deutsches Ärzteblatt*, 107(4), 50–56.
- Wiegand, J. (2011). *Simulatorgestütztes Verkehrssicherheitstraining von Einsatzkräften*. Informationsblatt des Instituts für Arbeit und Gesundheit der Deutschen Gesetzlichen Unfallversicherung, Ausgabe 12/2011, Nr. 3046.

- Wiegand, J. & Pieper-Nagel, M. (2011). Zertifizierte Fortbildung für mehr Verkehrssicherheit: Training für Fahrer von Einsatzfahrzeugen. *Rettungsdienst*, 34, 42-46.
- Wilde, G. J. S. (1998). Risk homeostasis theory: an overview. *Injury Prevention*, 4, 89-91.
- Wittchen, H.-U. & Jacobi, F. (2007). *Brauchen wir neue Perspektiven in der Wirkstoffforschung der Psychotherapie?* (S. 1-45). In Becker, R. & Wunderlich, H.-P. (Hrsg.). *Wie wirkt Psychotherapie?: Forschungsgrundlagen für die Praxis*. Stuttgart: Thieme.
- Wittchen, H.-U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B. et al. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21, 655–679.

## Abbildungsverzeichnis

Abbildung 1: Neues Kirkpatrick Evaluationsmodell („New World Kirkpatrick Model“, Kirkpatrick & Kirkpatrick, 2016, übersetzt durch Autoren).....	8
Abbildung 2: Datenerhebung – Ablauf für Interventionsgruppen mit Vorhermessung (nur T1/T2-Messung) .....	15
Abbildung 3: Überblick der erhobenen Daten mit Angabe der Abschnitte, in denen die Methoden beschrieben werden .....	16
Abbildung 4: Teilnehmer und Dropout-Übersicht nach CONSORT-Schema .....	34
Abbildung 5: Datenmodell der Studie.....	36
Abbildung 6: Auswahl kategorialer demographischer Daten (N=183 Probanden) .....	38
Abbildung 7: Personen mit potenziellen Deckeneffekten zum ersten Messzeitpunkt.....	46
Abbildung 8: Zusammensetzung der durchgeführten Trainings .....	50
Abbildung 9: Bewertung der Trainingsinhalte in Abhängigkeit von Versuchsgruppe und Alters der Probanden .....	55
Abbildung 10: positive Kommentare direkt nach dem Training in Abhängigkeit von der Versuchsgruppe .....	58
Abbildung 11: negative Kommentare direkt nach dem Training in Abhängigkeit von der Versuchsgruppe .....	59
Abbildung 12: Freie Antworten 1-2 Monate nach dem Training in Abhängigkeit von der Versuchsgruppe .....	60
Abbildung 13: Freie Antworten 6 Monate nach dem Training in Abhängigkeit von der Versuchsgruppe .....	61
Abbildung 14: Erreichte Gesamtpunkte im Wissenstest im zeitlichen Verlauf .....	64
Abbildung 15: Wissenstestbereiche mit signifikanten Interaktionseffekten zwischen den Gruppen und Zeitpunkten .....	65
Abbildung 16: Erreichte Punktzahl im Wissenstest unterteilt nach Versuchsgruppe und Alter der Probanden .....	66
Abbildung 17: Interaktionseffekt zwischen Gruppen und Zeitpunkten bei verkehrsbezogener Extraversion .....	66
Abbildung 18: Altersabhängigkeit verkehrsbezogener Extraversion .....	67
Abbildung 19: Interaktionseffekt zwischen Geschlecht und Zeitpunkten auf verkehrsbezogene Gewissenhaftigkeit .....	68
Abbildung 20: Trainingseffekt der Risikowahrnehmung bei Normalfahrten.....	69
Abbildung 21: Risk Seeking im zeitlichen Verlauf in Abhängigkeit vom Geschlecht .....	70
Abbildung 22: Fahrfähigkeit (alle Skalen) im zeitlichen Verlauf .....	71
Abbildung 23: Verlauf der wahrgenommenen eigenen Unfallwahrscheinlichkeit in Abhängigkeit von Alter und Geschlecht .....	72
Abbildung 24: Mittlere und maximale Geschwindigkeit der RTW-Fahrten zum Einsatzort kontrolliert für Alter und Geschlecht.....	75
Abbildung 25: Maximale Geschwindigkeit bei RTW-Fahrten zum Einsatzort in Abhängigkeit vom Geschlecht .....	75
Abbildung 26: Längsbeschleunigungen bei RTW-Fahrten zum Einsatzort in Abhängigkeit von Geschlecht und Gruppe .....	77
Abbildung 27: Querschleunigungen bei RTW-Fahrten zum Einsatzort in Abhängigkeit von Geschlecht und Gruppe .....	78
Abbildung 28: Beifahrerbeurteilung vorausschauender Fahrweise in Abhängigkeit vom Alter des Fahrers .....	79
Abbildung 29: Beifahrerbeurteilung rasanter Fahrweise in Abhängigkeit vom Alter des Fahrers ...	80
Abbildung 30: Prozenträge für kognitive und emotionale Irritation.....	84

Abbildung 31: Erschöpfungsneigung und Magenbeschwerden der Studiengruppen im Vergleich zur Allgemeinbevölkerung .....	85
Abbildung 32: Gliederschmerzen und Herzbeschwerden der Studiengruppen im Vergleich zur Allgemeinbevölkerung .....	85
Abbildung 33: Zynismus (Burnout) in Abhängigkeit von Alter und Geschlecht .....	86
Abbildung 34: Depressionsscreening, ca. 34% der Personen mit Depressionsverdacht weisen tatsächlich eine Depression auf.....	86
Abbildung 35: Mittlere Veränderung des intrapsychischen Gleichgewichtszustands über die Rettungsdienstschichten nach Geschlecht .....	87
Abbildung 36: Geistige und körperliche Anforderungen zu Schichtende .....	87
Abbildung 37: Frustration zu Schichtende in Abhängigkeit vom Geschlecht .....	88
Abbildung 38: mittlere (durchgezogene Linie), maximale und minimale (gepunktete Linie) Herzfrequenz in den Versuchsgruppen in den Schichtmessungen.....	89
Abbildung 39: Veranschaulichung der Zusammenführung der Messzeitpunkte .....	92
Abbildung 40: Risikowahrnehmung für mittelschwere und leichte Verstöße sowie bei Normalfahrten in der Wartegruppe in Abhängigkeit vom Alter .....	97
Abbildung 41: Risikowahrnehmung bei Normalfahrten in der ungepaarten Stichprobe .....	100
Abbildung 42: Mittlere Geschwindigkeiten bei RTW-Fahrten zum Einsatzort im Zeitverlauf in der Wartegruppe .....	104
Abbildung 43: Selbst- und Fremdbeurteilung der sicheren Fahrweise im zeitlichen Verlauf für die Wartegruppe.....	104
Abbildung 44: Mittlere und maximale Geschwindigkeit bei RTW-Fahrten zum Einsatzort im kurzfristigen Verlauf unter Hinzunahme der Dropouts .....	108
Abbildung 45: Mittlere und maximale Geschwindigkeit bei RTW-Fahrten zum Einsatzort im langfristigen Vergleich unter Hinzunahme der Dropouts .....	109
Abbildung 46: Kognitive Irritation in Abhängigkeit vom Geschlecht .....	112
Abbildung 47: Geistige Anforderungen, Anstrengung und Frustration zum Schichtende im zeitlichen Verlauf für die Wartegruppe.....	113
Abbildung 48: Kognitive und emotionale Irritation im langfristigen Verlauf unterteilt nach Versuchsgruppen .....	115
Abbildung 49: Geistige Anforderungen zum Schichtende in Abhängigkeit vom Alter im langzeitlichen Verlauf .....	116
Abbildung 50: Verteilung der Kategorien potenziell kritischer Ereignisse vor dem Training.....	163
Abbildung 51: Ortslagen der potenziell kritischen Ereignisse in Abhängigkeit der Ereigniskategorien .....	164
Abbildung 52: Straßengröße in Fahrtrichtung vor oder während des potenziell kritischen Ereignisses in Abhängigkeit der Ereigniskategorie vor dem Training.....	164
Abbildung 53: Witterungs- und Sichtverhältnisse während der potenziell kritischen Ereignisse in Abhängigkeit der Ereigniskategorie .....	165
Abbildung 54: Vorfahrt: Arten und Verkehrsdichte vor den potenziell kritischen Situationen .....	166
Abbildung 55: Vorfahrtssituationen: Ausprägungen des Quer- und Gegenverkehrs.....	167
Abbildung 56: Arten von Überholvorgängen.....	168
Abbildung 57: Kodierung des Gegenverkehrs bei Überholvorgängen .....	169
Abbildung 58: Arten von sonstigen potenziell kritischen Ereignissen .....	170
Abbildung 59: Ausprägung des Quer- und Gegenverkehrs bei den sonstigen potenziell kritischen Ereignissen .....	170

## Tabellenverzeichnis

Tabelle 1: Versuchsplan nach der Anpassung .....	16
Tabelle 2: Interne Konsistenzen der Trainingsevaluationsbögen .....	17
Tabelle 3: Neue Skalenstruktur der verkehrsbezogenen Persönlichkeitseigenschaften .....	18
Tabelle 4: Neue Skalenstruktur der verkehrsbezogenen Risikowahrnehmung .....	19
Tabelle 5: Kategoriensystem zur Beurteilung der freien Antworten im Wissenstest .....	24
Tabelle 6: Wissenstestbereiche mit Beispielitems .....	25
Tabelle 7: Verteilung der Wissenstestversionen innerhalb der Versuchsgruppen .....	26
Tabelle 8: Organisationsakquise aller in Frage kommenden Rettungsdienste .....	33
Tabelle 9: Zellbesetzungen auf jeder Evaluations- und zeitlichen Ebene .....	36
Tabelle 10: Demographische Daten der Gesamtstichprobe zu Studienbeginn.....	37
Tabelle 11: Unterschiede in den (berufsbezogenen) demographischen Merkmalen aller aufgenommenen Teilnehmer zwischen den Gruppen zu Studienbeginn .....	40
Tabelle 12: Unterschiede in den verkehrsrelevanten Merkmalen aller aufgenommenen Teilnehmer zwischen den Gruppen zu Studienbeginn.....	41
Tabelle 13: Unterschiede in den medizinischen Merkmalen aller aufgenommenen Teilnehmer zwischen den Gruppen zu Studienbeginn .....	44
Tabelle 14: Teilnehmer mit guten Fahrprofilen zu Studienbeginn im Vergleich zu anderen Teilnehmern: Demographische Aspekte .....	47
Tabelle 15: Teilnehmer mit guten Fahrprofilen zu Studienbeginn im Vergleich zu anderen Teilnehmern: Arbeitsbedingungen.....	49
Tabelle 16: Überblick über die abhängigen Variablen zum Zeitpunkt T1 über alle Gruppen .....	51
Tabelle 17: Haupteffekte der Evaluationsebene „Reaktion“ .....	54
Tabelle 18: Unterschiede in der Trainingsbewertung in Abhängigkeit vom Alter.....	55
Tabelle 19: Unterschiede in den Reaktionen zwischen den Orten der Trainingsdurchführung....	56
Tabelle 20: Zeitlicher Verlauf in der Trainingsbewertung direkt nach dem Training sowie 1-2 Monate später .....	57
Tabelle 21: Langfristiger zeitlicher Verlauf in der Trainingsbewertung direkt nach dem Training, 1-2 Monate später sowie 6 Monate später .....	57
Tabelle 22: Haupteffekte der Evaluationsebene „Lernen“ .....	62
Tabelle 23: Unterschiede in den verschiedenen Wissenstestbereichen .....	64
Tabelle 24: Alterseinfluss auf verkehrsbezogene Persönlichkeitseigenschaften.....	67
Tabelle 25: Einfluss des Geschlechts bei verkehrsbezogenen Persönlichkeitseigenschaften.....	68
Tabelle 26: Alterseinfluss auf Risikobewusstsein .....	69
Tabelle 27: Haupteffekte der Evaluationsebene „Verhalten“ .....	73
Tabelle 28: Einfluss des Geschlechts auf die Längsbeschleunigungen .....	76
Tabelle 29: Einfluss des Geschlechts auf die Querschleunigungen.....	79
Tabelle 30: Haupteffekte der Evaluationsebene „Resultate“ .....	81
Tabelle 31: Kontrolle der Messeffekte: Unterschiede zwischen Gruppe 1 und 3 zum Zeitpunkt T2.....	90
Tabelle 32: Bestimmung der Voraussetzungen zur Verschiebung der Messzeitpunkte der Gruppe 2.....	93
Tabelle 33: Lernebene: Zeitlicher Verlauf über die ersten drei Messzeitpunkte in der Wartegruppe .....	96
Tabelle 34: Lernebene: Langfristiger Verlauf in der Gesamtstichprobe .....	98
Tabelle 35: Lernebene: Kurzfristige Effekte in der ungepaarten Stichprobe .....	99
Tabelle 36: Lernebene: Langfristige Effekte in der ungepaarten Stichprobe.....	101
Tabelle 37: Verhaltensebene: Zeitlicher Verlauf über die ersten drei Messzeitpunkte in der Wartegruppe .....	102
Tabelle 38: Verhaltensebene: Langfristiger Verlauf in der Gesamtstichprobe.....	105
Tabelle 39: Verhaltensebene: Kurzfristige Effekte in der ungepaarten Stichprobe.....	107



Tabelle 40:	Verhaltensebene: Langfristige Effekte in der ungepaarten Stichprobe .....	109
Tabelle 41:	Resultateebene: Zeitlicher Verlauf über die ersten drei Messzeitpunkte in der Wartegruppe .....	110
Tabelle 42:	Resultateebene: Langfristiger Verlauf in der Gesamtstichprobe.....	114
Tabelle 43:	Resultateebene: Kurzfristige Effekte in der ungepaarten Stichprobe.....	117
Tabelle 44:	Resultateebene: Langfristige Effekte in der ungepaarten Stichprobe .....	118
Tabelle 45:	Regression zur Kontrolle des Erfahrungseinflusses auf die kurzfristigen Effekte des Trainings (I) .....	121
Tabelle 46:	Regression zur Kontrolle des Erfahrungseinflusses auf die kurzfristigen Effekte des Trainings (II) .....	121
Tabelle 47:	Regression zur Kontrolle des Erfahrungseinflusses für die langfristigen Effekte des Trainings (I) .....	122
Tabelle 48:	Regression zur Kontrolle des Erfahrungseinflusses für die langfristigen Effekte des Trainings (II) .....	122
Tabelle 49:	Regression zur Kontrolle des Erfahrungseinflusses für die langfristigen Effekte des Trainings (III) .....	123
Tabelle 50:	Prozentuale Verteilung des Trainingserfolgs in einzelnen Merkmalen.....	126
Tabelle 51:	Prozentuale Verteilung des Trainingserfolgs auf den verschiedenen Evaluationsebenen und global .....	128
Tabelle 52:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Lernebene durch individuelle Voraussetzungen .....	129
Tabelle 53:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Verhaltensebene durch individuelle Voraussetzungen.....	131
Tabelle 54:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Resultateebene durch individuelle Voraussetzungen.....	134
Tabelle 55:	Prädiktion kurzfristigen und längerfristigen Gesamterfolgs durch individuelle Voraussetzungen .....	136
Tabelle 56:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Lernebene durch organisationsbezogene Voraussetzungen.....	139
Tabelle 57:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Verhaltensebene durch organisationsbezogene Voraussetzungen.....	140
Tabelle 58:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Resultateebene durch organisationsbezogene Voraussetzungen.....	142
Tabelle 59:	Prädiktion kurzfristigen und längerfristigen Gesamterfolgs durch organisationsbezogene Voraussetzungen.....	144
Tabelle 60:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Lernebene durch gruppenbezogene Voraussetzungen.....	147
Tabelle 61:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Verhaltensebene durch gruppenbezogene Voraussetzungen.....	149
Tabelle 62:	Prädiktion kurzfristigen und längerfristigen Erfolgs auf Resultateebene durch gruppenbezogene Voraussetzungen.....	151
Tabelle 63:	Prädiktion kurzfristigen und längerfristigen Gesamterfolgs durch gruppenbezogene Voraussetzungen .....	153
Tabelle 64:	Unterschiede zwischen RTW- Sondersignalfahrten zum Einsatzort und Sondersignalfahrten mit Patient .....	157
Tabelle 65:	Korrelationen zwischen den abhängigen Variablen der RTW-Sondersignalfahrten und schichtbezogenen Kontrollvariablen .....	159
Tabelle 66:	Häufigkeiten der Reaktionen auf die Arten der Vorfahrtssituationen .....	167
Tabelle 67:	Häufigkeiten der Reaktionen auf die Arten der Überholsituationen .....	169
Tabelle 68:	Häufigkeiten der Reaktionen auf die Arten der sonstigen Situationen .....	171
Tabelle 69:	Einflussfaktoren auf die EKG-Messungen zum Zeitpunkt T1.....	172

## Anhang/Anhänge

## Anhang A

### Einflussfaktoren auf Einsatzfahrten

Die Berechnung für alle folgenden Tabellen dieses Abschnitts erfolgte mittels hierarchischen Regressionen auf die Veränderung der abhängigen Variablen ( $\Delta$ : Differenz Zeitpunkte T2 – T1). Im ersten Schritt wird das Alter und Geschlecht aufgenommen. In Schritt zwei wurden erfahrungsspezifische Variablen (Qualifikation, Rettungsdienst Erfahrung, Fahrzeiten im Rettungsdienst, Anzahl der Führerscheinklassen und die Verkehrserfahrung) schrittweise hinzugefügt. Im dritten Schritt wurden schrittweise die Einsatzschwere, Art der Schicht, Schichtdauer, Anzahl der Einsätze, Anzahl selbstberichteter kritischer Ereignisse, Müdigkeit sowie zeitliche und geistige Anforderungen über die Schichten einberechnet. In den Tabellen sind für den zweiten und dritten Schritt jeweils nur die Variablen aufgeführt, die einen Einfluss auf die unabhängige Variable zeigten.

Abhängige Variable	$\Delta$ Mittlere Geschwindigkeit Hinfahrt (N=96)			$\Delta$ Mittlere Geschwindigkeit Patiententransport (N=80)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.019	.019		.020	.020
...Alter	0.13			0.14		
...Geschlecht	0.03			-0.05		
Schritt 2: Erfahrung		.104*	.085**			
...Führerscheinklassen	0.32**					
Schritt 3: Schichtdaten		.251***	.147***		.077*	.058*
...Anzahl Einsätze	-0.41			-0.25*		
...zeitliche Anforderungen	0.21*					
Abhängige Variable	$\Delta$ Maximale Geschwindigkeit Hinfahrt (N=96)			$\Delta$ Maximale Geschwindigkeit Patiententransport (N=80)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.016	.016		.018	.018
...Alter	-0.05			0.12		
...Geschlecht	0.13			-0.10		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.076 <sup>+</sup>	.060*		.101*	.083*
...Anzahl Einsätze	0.25*					
...Schichtart				-0.29**		

Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Abhängige Variable	Δ Maximale Längsbeschleunigung Hinfahrt (N=96)			Δ Maximale Längsbeschleunigung Patiententransport (N=80)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.036	.036		.030	.030
...Alter	-0.01			-0.05		
...Geschlecht	0.19			0.18		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.150**	.114**		.268***	.238**
...Schichtart				-0.36***		
...geistige Anforderungen	-0.34**			0.35**		
Abhängige Variable	Δ Mittlere Längsbeschleunigung Hinfahrt (N=96)			Δ Mittlere Längsbeschleunigung Patiententransport (N=80)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.048 <sup>+</sup>	.048*		.013	.013
...Alter	-0.09			-0.05		
...Geschlecht	0.22*			0.12		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.321***	.273***		.252**	.239**
...Schichtart	-0.42***			-0.36**		
...Anzahl Einsätze	-0.34***			-0.29**		
...geistige Anforderungen				0.25*		
Abhängige Variable	Δ Maximale Bremsverzögerung Hinfahrt (N=96)			Δ Maximale Bremsverzögerung Patiententransport (N=80)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.083*	.083*		.018	.018
...Alter	0.13			0.08		
...Geschlecht	-0.29**			-0.13		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.141**	.058*		.320***	.302***
...Schichtart	0.24*			0.38***		
...geistige Anforderungen				-0.43***		

Abkürzungen: N=Anzahl der Probanden; β=standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung, ΔR<sup>2</sup>=Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Abhängige Variable	Δ Mittlere Bremsverzögerung Hinfahrt (N=96)			Δ Mittlere Bremsverzögerung Patiententransport (N=80)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.050 <sup>+</sup>	.050 <sup>+</sup>		.009	.009
...Alter	0.15			0.01		
...Geschlecht	-0.21 <sup>*</sup>			-0.10		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.329 <sup>***</sup>	.279 <sup>***</sup>		.323 <sup>***</sup>	.314 <sup>***</sup>
...Schichtart	0.49 <sup>***</sup>			0.41 <sup>***</sup>		
...Anzahl Einsätze	0.24 <sup>**</sup>					
...Ermüdung				0.28 <sup>**</sup>		
...geistige Anforderungen				-0.21 <sup>*</sup>		
Abhängige Variable	Δ Mittlere positive Querbearbeitung Hinfahrt (N=96)			Δ Mittlere positive Querbearbeitung Patiententransport (N=80)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.012	.012		.004	.004
...Alter	-0.10			-0.05		
...Geschlecht	0.08			0.05		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.214 <sup>***</sup>	.202 <sup>***</sup>		.255 <sup>***</sup>	.255 <sup>**</sup>
...Anzahl Einsätze	-0.22 <sup>*</sup>					
...Schichtart	-0.41 <sup>***</sup>			-0.30 <sup>**</sup>		
...geistige Anforderungen				0.32 <sup>**</sup>		
...Ermüdung				-0.25 <sup>*</sup>		
Abhängige Variable	Δ Mittlere negative Querbearbeitung Hinfahrt (N=96)			Δ Mittlere negative Querbearbeitung Patiententransport (N=80)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.014	.014		.013	.013
...Alter	0.10			0.11		
...Geschlecht	-0.09			-0.07		
Schritt 2: Erfahrung						
Schritt 3: Schichtdaten		.247 <sup>***</sup>	.213 <sup>***</sup>		.278 <sup>***</sup>	.265 <sup>**</sup>
...Anzahl Einsätze	0.28 <sup>**</sup>					
...Schichtart	0.41 <sup>***</sup>			0.31 <sup>**</sup>		
...geistige Anforderungen				-0.33 <sup>**</sup>		
...Ermüdung				0.25 <sup>*</sup>		

Abkürzungen: N=Anzahl der Probanden; β=standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung, ΔR<sup>2</sup>=Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

## Anhang B

### Einflussfaktoren auf Beanspruchung

Die Berechnung für alle folgenden Tabellen dieses Abschnitts erfolgte mittels schrittweiser Regression unter Einschluss des Schritt 1 zum Zeitpunkt T1. Im ersten Schritt wird das Alter und Geschlecht aufgenommen. Im zweiten Schritt wurden gesundheitsbezogenen Variablen (Raucherstatus, sportliche Aktivität, BMI, Herzkreislauferkrankungen, Medikamente) aufgenommen. In Schritt drei wurden erfahrungsspezifische Variablen (Qualifikation, Rettungsdienst Erfahrung, Fahrzeiten im Rettungsdienst und Führerscheinbesitz) schrittweise hinzugefügt. Im dritten Schritt wurden schrittweise der Schichtbeginn, die Schichtdauer und –art sowie die Anzahl der Einsätze einberechnet. In den Tabellen werden für den zweiten bis vierten Schritt jeweils nur die Variablen aufgeführt, die einen Einfluss auf die unabhängige Variable zeigten.

Abhängige Variable	Mittlere Müdigkeit (N=106)			Ermüdung (N=97)			Veränderung der Vitalität (N=99)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.030	.030		.001	.001		.021	.021
...Alter	-0.08			-0.02			0.10		
...Geschlecht	-0.13			0.04			0.08		
Schritt 2: Gesundheit									
Schritt 3: Erfahrung		.089*	.059*						
...Qualifikation	0.25*								
Schritt 4: Schichtdaten		.178***	.090**		.247***	.246***		.193***	.171**
...Schichtart	0.31**								
...Schichtbeginn				0.47***			-0.33**		
...Anzahl Einsätze				0.21*			-0.27**		
Abhängige Variable	Veränderung des Gleichgewichtszustandes (N=99)			Veränderung der Exrtvertiertheit (N=99)			Veränderung der Vitalität (N=99)		
	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$	$\beta$	R <sup>2</sup>	$\Delta R^2$
Schritt 1: Demographie		.041	.041		.014	.014		.014	.014
...Alter	-0.17			0.10			-0.05		
...Geschlecht	-0.06			0.04			0.13		
Schritt 2: Gesundheit					.068 <sup>+</sup>	.054*			
...körperliche Aktivität				-0.32*					
Schritt 3: Erfahrung									
Schritt 4: Schichtdaten		.176***	.169**		.182**	.114***		.199***	.185***
...Schichtbeginn	-0.27**			-0.35***			-0.44***		
...Anzahl Einsätze	-0.32**								

Abkürzungen: N=Anzahl der Probanden;  $\beta$ =standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung,  $\Delta R^2$ =Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001

Abhängige Variable	Geistige Anforderungen (N=105)			Körperliche Anforderungen (N=105)			Zeitliche Anforderungen (N=105)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.028	.028		.009	.009		.045 <sup>+</sup>	.045 <sup>+</sup>
...Alter	0.17			0.10			0.20*		
...Geschlecht	-0.11			-0.02			0.02		
Schritt 2: Gesundheit								.110**	.065*
...Anzahl Medikamente							-0.27**		
Schritt 3: Erfahrung		.079*	.050*		.058	.048*		.170**	.060**
...Rettungsdienstmonate	0.39*			0.38*			0.43**		
Schritt 4: Schichtdaten		.195**	.116**		.215***	.157**		.258***	.089**
...Schichtdauer	0.27**			0.32**					
...Anzahl Einsätze	0.22*			0.24**			0.30**		
Abhängige Variable	Leistung (N=104)			Anstrengung (N=104)			Frustration (N=105)		
	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>	β	R <sup>2</sup>	ΔR <sup>2</sup>
Schritt 1: Demographie		.018	.018		.010	.010		.016	.016
...Alter	-0.03			0.11			-0.13		
...Geschlecht	0.14			-0.02			0.01		
Schritt 2: Gesundheit									
Schritt 3: Erfahrung					.088*	.077**			
...Rettungsdienstmonate				0.48**					
Schritt 4: Schichtdaten					.172**	.084**		.070*	.054*
...Schichtdauer				0.30**					
...Anzahl Einsätze							0.23*		

Abkürzungen: N=Anzahl der Probanden; β=standardisierter Beta-Koeffizient, R<sup>2</sup>=Varianzaufklärung, ΔR<sup>2</sup>=Änderung der Varianzaufklärung, Signifikanzniveau: +p ≤ .10; \*p ≤ .05; \*\*p ≤ .01; \*\*\*p ≤ .001