



universität  
wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Zur Verbesserung der Reliabilität im Reaktionszeitexperiment  
durch Fehler- und Ratekorrekturen“

verfasst von / submitted by

Anna-Maria Mayer, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt / degree programme  
code as it appears on  
the student record sheet:

A 066 840

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Masterstudium Psychologie UG2002

Betreut von/ Supervisor:

Univ.-Prof. Dr. rer. nat. habil. Matthias Gondan-Rochon

## **Abstract**

In einem 2AFC-Experiment wird die Testperson instruiert, so schnell und so genau wie möglich zu reagieren. Reagieren Testpersonen genau, reagieren sie langsamer als sie könnten. Reagieren Testpersonen schnell, machen sie eher fehlerhafte Reaktionen. Fehlreaktionen gehören genauso wie richtige Reaktionen zur Performanz von Testpersonen, werden allerdings oft aus der Analyse von z.B. der mittleren Reaktionszeit ausgeschlossen (entspricht dem Reduced-Sample Schätzer). Aus statistischer Sicht handelt es sich bei Fehlreaktionen um ein Zensierungsproblem. Die Kaplan-Meier Methode und eine adaptierte Form des Aalen-Johansen Schätzers können, angepasst an die Daten eines Reaktionszeitexperimentes, als Korrektur für Fehlreaktionen verwendet werden. Einige Fehlreaktionen sind auf Ratedurchgänge rückführbar. Diese Ratedurchgänge können mit der Kill-the-Twin Prozedur korrigiert werden. Es ist annehmbar, dass die theoriegeleitete Korrektur von Fehlreaktionen die „wahre“ Performanz einer Testperson besser abbildet und daher auch reliabler ist. Bei der Berechnung der Reliabilität sollte der Reduced-Sample Schätzer schlechter abschneiden, als die Fehlerkorrekturen alleine und diese wiederum schlechter als Fehler- mit Ratekorrekturen. In den vier verwendeten Reaktionszeittests ist diese Annahme tendenziell bestätigt. Unter Anwendung von Fehler- und Ratekorrekturen verbessert sich die Reliabilität in allen Verfahren geringfügig.

## **Abstract**

In a 2AFC-Experiment, participants are usually asked to respond as fast and as accurate as possible. This instruction results in a conflict of strategies: If a participant responds accurately, the velocity of the response decreases. If a participant responds fast, wrong reactions are made more frequently. Wrong reactions are often excluded from the analysis f.e. of average reaction times (= reduced-sample estimator). From a statistical point of view, a wrong reaction results in a censoring problem, which could be resolved with the Kaplan-Meier method or an adapted version of the Aalen-Johansen estimator. Wrong reactions are sometimes caused by guessing trials. If guessing trials are subtracted from the data, theoretically, the “true” performance should be estimated even more accurately. The Kaplan-Meier method, the Aalen-Johansen estimator and a combination of both with the Kill-the-Twin procedure should lead to more accurate, hence more reliable estimations of median reaction times. A tendency towards more reliable estimations for wrong-reaction-corrections in combination with guessing corrections was found, followed by more reliable estimations for wrong-reaction-corrections compared to the reduced-sample estimator.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>5</b>
<b>2</b>	<b>Reduced-Sample Schätzer (RS)</b>	<b>6</b>
2.1	Einschränkungen des RS .....	7
<b>3</b>	<b>Kaplan-Meier Methode (KM)</b>	<b>9</b>
3.1	Implementierung der KM in Reaktionszeitexperimenten .....	10
<b>4</b>	<b>Aalen-Johansen Schätzer (AJ)</b>	<b>12</b>
4.1	Implementierung des AJ in Reaktionszeitexperimenten .....	12
4.2	Kritik an sequentiellen-sampling Modellen .....	14
<b>5</b>	<b>Kill-the-Twin Prozedur (KTT)</b>	<b>14</b>
5.1	Exkurs: die Wettlaufungleichung.....	17
<b>6</b>	<b>Annahmen</b>	<b>18</b>
<b>7</b>	<b>Methode und Material</b>	<b>20</b>
7.1	visual search task.....	20
7.1.1	Implementierung der KTT im visual search task.....	21
7.2	divided attention .....	21
7.2.1	Implementierung der KTT in divided attention. ....	22
7.3	go-nogo.....	23
7.4	vigilance .....	24
7.5	Statistische Auswertung .....	25
7.5.1	Implementierung der Schätzer. ....	25
<b>8</b>	<b>Ergebnisse</b>	<b>25</b>
8.1	Retestrelabilität des visual search task .....	25
8.2	Retestrelabilität der divided attention .....	26
8.3	Split-Half-Reliabilität des go-nogo .....	27
8.4	Split-Half-Reliabilität der vigilance .....	28

<b>9 Diskussion</b>	<b>28</b>
<b>Referenzen</b>	<b>32</b>
<b>Abbildungsverzeichnis</b>	<b>35</b>
<b>Tabellenverzeichnis</b>	<b>37</b>
<b>Abkürzungsverzeichnis</b>	<b>37</b>

# 1 Einleitung

Bei der Untersuchung psychischer Phänomene spielt das Reaktionszeitexperiment eine wichtige Rolle. Im typischen Reaktionszeitexperiment bittet man Testpersonen, auf bestimmte Reize so schnell und so genau wie möglich zu antworten. Allerdings ist dieser konträren Anweisung schwer Folge zu leisten. Wird genau geantwortet, sinkt die Geschwindigkeit der Reaktionen, da die Testperson vorsichtiger ist. Wird so schnell wie möglich geantwortet, steigt die Fehlerrate, da mehr Risiko eingegangen wird. In typischen Reaktionszeitexperimenten werden Fehlreaktionen nicht berücksichtigt; zum Beispiel basiert die mittlere Reaktionszeit i.d.R. auf den korrekten Reaktionen. In der Terminologie der Überlebenszeitanalyse ist das der sog. Reduced-Sample Schätzer (RS; Kaplan & Meier, 1958). Diesem Vorgehen liegen zwei Annahmen zugrunde: erstens, dass konzentriertes Arbeiten immer zu korrekten Reaktionen führt. Zweitens, dass Fehlreaktionen keine Information über zugrundeliegende mentale Prozesse liefern (weil sich die Testperson nicht „konzentriert“ hat). Die erste Annahme ist nur für extrem leichte Aufgaben haltbar. Bei schwierigeren Aufgaben können durchaus Fehler passieren, auch wenn man sich konzentriert. In diesem Fall liefert eine Fehlreaktion zum Zeitpunkt  $t$  durchaus Information, nämlich, dass die Testperson bis zu diesem Zeitpunkt die Antwort nicht wusste. Hätte sie länger überlegt, hätte sie die korrekte Antwort später gegeben, oder gar nicht, was zu einer unendlich langen Reaktionszeit führen würde (was auch nach  $t$  liegt).

Aus statistischer Sicht handelt es sich dabei um ein Zensierungsproblem; die Fehlreaktion zensiert die korrekte Reaktion. Sind diese beiden Prozesse unabhängig, liegt eine uninformative Zensierung vor, dann liefert die Kaplan-Meier Methode (KM; Kaplan & Meier, 1958) eine erwartungstreue Schätzung der Zeit, die die Testperson gebraucht hätte um korrekt zu reagieren. Sind Fehlreaktionen und korrekte Reaktionen abhängig voneinander (liegen sie z.B. auf den entgegengesetzten Enden einer Evidenzdimension, Ratcliff, 1978), können die Zeiten für Fehlreaktionen mit einer adaptierten Form des Aalen-Johansen Schätzers (AJ; Aalen & Johansen, 1978; Borgan, 2015) berechnet werden.

Im 2AFC-Task hat eine Testperson eine 50 : 50 Chance richtig zu reagieren, wenn sie rät; daher kann man davon ausgehen, dass auf jeden Fehler auch eine richtig erratene korrekte Reaktion kommt. Um den Anteil an Rateversuchen in die Auswertung miteinzubeziehen, kann die Kill-the-Twin Prozedur (KTT; Eriksen, 1988) mit den Schätzern kombiniert werden. Die Funktionsweise der einzelnen Schätzer soll am Beispiel aus Tabelle 1 illustriert werden. An einer Gegenüberstellung der einzelnen Methoden soll gezeigt werden, wie sich die Reaktionszeit von Testperson 2 ändert.

Tabelle 1 *Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit \*)*.

Trial	Testperson 1	Testperson 2
1	200	200
2	210	210
3	220	220
4	230	230
5	240	240
6	250	250
7	260	260
8	270	270
9	280	280*
10	290	290
11	300	300
	$\bar{x} = 250$	$\bar{x} = ?$

## 2 Reduced-Sample Schätzer (RS)

Wie in Kap. 1 ausgeführt, werden bei Reaktionszeitexperimenten oftmals nur die korrekten Reaktionen beachtet. Fehlerhaftes Verhalten wie falsche Reaktionen, Auslassungen, „Ausreißer“ usw. werden a priori ausgeschlossen (siehe auch: Gondan und Minakata, 2016) und bleiben bei der Berechnung der mittleren Reaktionszeit unberücksichtigt. Diesem Umgang mit Fehlreaktionen liegen die Annahmen zugrunde, dass konzentriertes Arbeiten immer zu korrekten Reaktionen führt und, dass Fehlreaktionen keine Information über zugrundeliegende mentale Prozesse liefern. Es sind keine Rückschlüsse über die „wahre“ Dauer einer Fehlreaktion möglich, da diese uninformativ ist.

Diese Annahmen werden ohne empirische Evidenz als gültig angenommen. Dabei bleibt unklar, ob es der RS ein adäquater Umgang mit Fehlreaktionen ist. Die Frage nach dem adäquaten Umgang von Fehlreaktionen muss allerdings auf Basis einer theoretischen und empirischen Fundierung gestellt werden. Schließlich gibt es kaum eine Testperson, die innerhalb eines Reaktionszeitexperimentes keine Fehler macht. Die Fehlerrate schwankt über Testpersonen hinweg, beträgt selten allerdings 0%.

Fehlerraten sind u.a. auf die Instruktionen eines klassischen Reaktionszeitexperimentes zurückzuführen: Testpersonen werden angewiesen so schnell und so genau wie möglich zu antworten. Geschwindigkeit und Genauigkeit stehen in einem Abhängigkeitsverhältnis zueinander. Dabei gilt generell: bei steigender Geschwindigkeit sinkt die Genauigkeit - bis hin zur reinen Zufallswahl der Reaktion. Ebenso gilt umgekehrt: bei steigender Genauigkeit sinkt die Geschwindigkeit (Pachella, 1974). Testpersonen müssen sich daher für ein Verhältnis Geschwindigkeit und Genauigkeit entscheiden (siehe Abbildung 1).

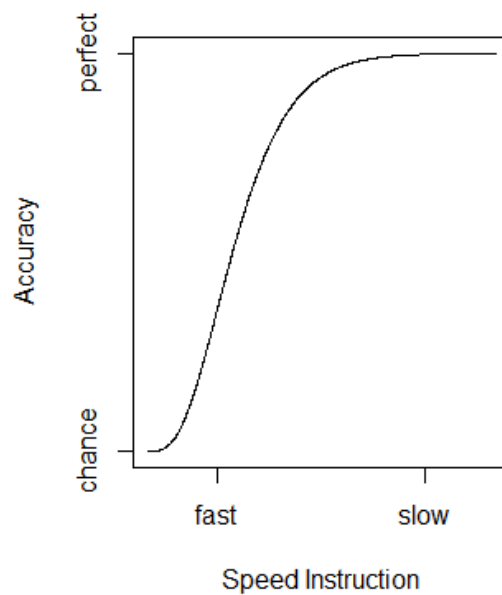


Abbildung 1 zeigt das theoretische Verhältnis von Geschwindigkeit zu Genauigkeit. Bei steigender Geschwindigkeit der mittleren Reaktionen sinkt die Genauigkeit auf Zufallsniveau herab. Bei steigender Genauigkeit verlangsamt sich die Geschwindigkeit der mittleren Reaktionen. Die Kurve nähert sich einer 100% Genauigkeit asymptotisch an, d.h. eine höhere Genauigkeit kann nur durch starke Reduktion der Geschwindigkeit erreicht werden.

## 2.1 Einschränkungen des RS

Die erste Annahme des RS („Konzentration“ führt zu richtigen Reaktionen) ist nur für extrem leichte Aufgaben haltbar. Bei schwierigeren Aufgaben können durchaus Fehler passieren, auch wenn man sich konzentriert. In diesem Fall liefert eine Fehlreaktion zum Zeitpunkt  $t$  die Information, dass die Testperson bis zu diesem Zeitpunkt die Antwort nicht wusste. Hätte sie länger überlegt, hätte sie die korrekte Antwort später gegeben, oder gar nicht, dann wäre die Reaktionszeit unendlich lang (was auch nach  $t$  liegt). Eine Fehlreaktion kann dieser Überlegung nach nicht mehr als uninformativ gelten. Sie informiert darüber, dass die „wahre“ Reaktionszeit nach  $t$  ms liegt. Solange diese Überlegung der Informativität von Fehlreaktionen gilt, führt die Schätzung von Reaktionszeitmaßen mit RS zu Verzerrungen. Insbesondere dann, wenn Fehlreaktionen nicht gleichmäßig in den Reaktionszeitdaten verteilt sind.

In Tabelle 1 hat Testperson 1 immer richtig reagiert. Testperson 2 hat bei 280 ms eine Fehlreaktion gemacht. Nach den vorangegangenen Überlegungen über die Informativität von Fehlreaktionen sollte die „wahre“ mittlere Reaktionszeit von Testperson 2 theoretisch langsamer sein. Wird die Fehlreaktion lediglich aus der Analyse exkludiert, ist jedoch nun Testperson 2 schneller als

Testperson 1 (Tab. 2). Die Performanz von Testperson 2 ist mit Anwendung des RS besser im Vergleich zu Testperson 1.

Tabelle 2. *Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktionen mit \*)*

Trial	Testperson 1	Testperson 2	Korrektur RS
1	200	200	200
2	210	210	210
3	220	220	220
4	230	230	230
5	240	240	240
6	250	250	250
7	260	260	260
8	270	270	270
9	280	280*	<del>280</del>
10	290	290	290
11	300	300	300
	$\bar{x} = 250$	$\bar{x} = ?$	$\bar{x} = 247$

Auch die Erkennungsrate (= Anteil der erkannten Zielreize von allen Zielreizen) wird verzerrt. Denn unter Ausschluss aller Fehlreaktionen beträgt die Erkennungsrate immer 100%, was eine Überschätzung der Performanz darstellt.

Eine Testperson reagiert in einem Reaktionszeitexperiment mit 100 Durchgängen 10-mal unter 300 ms, 10-mal über 300 ms und 80-mal falsch. Stellen sich die Forschenden die Frage, wie oft ein Zielreiz unter 300 ms erkannt wird, ergäbe sich mit RS eine Erkennungsrate von 50%. In der Hälfte aller Fälle würde die Testperson unter 300 ms antworten. Folglich ist die Erkennungsrate in diesem Fall überschätzt. Nur 5% aller Reaktionen waren schneller als 300 ms. 5% der Reaktionen waren langsamer als 300 ms und bei 90% der Durchgänge wurde kein Stimulus erkannt.

Der RS führt zu einer Verzerrung von Reaktionszeitmaßen, wie der mittleren Reaktionszeit und der Erkennungsrate. Zusätzlich ist die erste Annahme des RS nicht zwingenderweise richtig. Richtige Reaktionen müssen nicht nur auf „konzentriertes Arbeiten“ rückführbar sein, sondern könnten auch auf Raten, oder *deadlining* basieren. Deadlining ist besonders bei fixen Reaktionszeitintervallen beobachtbar. Eine Testperson reagiert nicht mehr so schnell, wie sie kann, sondern dem ihr zur Verfügung stehendem Intervall entsprechend. Hat eine Testperson 200 ms zu reagieren, werden sich ihre Reaktionszeiten diesem Limit annähern um einen passenden Speed-Accuracy Trade-Off zu erreichen.



Der RS stellt einen nicht zufriedenstellenden Schätzer der Performanz in Reaktionszeitexperimenten dar. Wie nun mit fehlerhaften Reaktionen noch umgegangen werden kann, soll in den Kapiteln 3–5 veranschaulicht werden.

### 3 Kaplan-Meier Methode (KM)

Die Kaplan-Meier Methode, oder auch der Produkt-Limit Schätzer genannt, ist eine Methode, die vor allem in klinischen Studien, wie Überlebenszeitanalysen angewandt wird. Mit ihr kann man Wahrscheinlichkeiten für ein Ereignis  $x$  (zumeist *Death*) in Abhängigkeit von der Zeit ( $t$ ) berechnen. Die Wahrscheinlichkeiten für  $x$  können auch bei zensierten Daten (*Loss*), beispielsweise bei Ausfällen oder wenn  $x$  nach dem Beobachtungszeitraum eintritt, berechnet werden (Ziegler, Lange & Bender, 2002), sofern *Death* und *Loss* als unabhängig voneinander gelten.

Bei der KM werden die Beobachtungszeiten durch das Eintreffen von  $x$  definiert und nicht vorab festgelegt (Ziegler et al., 2002). Für diese Zeiträume können bedingte Wahrscheinlichkeiten für das Eintreten von  $x$  berechnet werden (Kaplan & Meier, 1958). Die Gesamtwahrscheinlichkeit für  $x$  zum Zeitpunkt  $t$  ergibt sich aus dem Produkt der vorangegangenen bedingten Wahrscheinlichkeiten. Zensierte Daten werden bis zum Ende des Intervalls, in dem sie aufgetreten sind, bis zum nächsten  $x$ , als vorhanden angesehen (Ziegler et al., 2002). Für den Zeitraum bis zum nächsten Todesfall gelten sie als lebend, danach als verstorben.

Ein *Loss* gilt als uninformativ, d.h. diese PatientInnen sind weder systematisch kränker, noch gesünder als andere PatientInnen. Die Wahrscheinlichkeit der PatientInnen in einem späteren Intervall zu versterben, ist in jedem Intervall gleich wahrscheinlich. Die Wahrscheinlichkeit eines Losses zu versterben wird gleichmäßig nach rechts redistribuiert (Efron, 1967).

In einer Follow-Up Studie gibt es 10 Personen. 5 Personen versterben mit  $t = 1, 2, 3, 4, 5$  Wochen. Eine sechste Person zieht nach sechs Wochen in eine andere Stadt und kann nicht mehr beobachtet werden. Personen 7, 8, 9, 10 versterben mit  $t = 7, 8, 9, 10$  Wochen. Die sechste Person ist zensiert und es lässt sich nicht mit Sicherheit bestimmen, wann sie verstorben wäre. Für die Berechnung der bedingten Wahrscheinlichkeiten per Intervall bleibt die Zensierung bis zur 7. Woche erhalten. Danach wird die Wahrscheinlichkeit der sechsten Person zu versterben gleichmäßig nach rechts redistribuiert. Bis zur 7. Woche beträgt die Wahrscheinlichkeit zu versterben je 10% pro Intervall. In den nachfolgenden Intervallen steigt die Wahrscheinlichkeit auf je 13% pro Intervall.

### 3.1 Implementierung der KM in Reaktionszeitexperimenten

Wettlaufmodelle sind eine Untergruppe der sequentiellen-sampling Modelle. Alle sequentiellen-sampling Modelle haben die Annahme, dass Information vor einer Reaktion akkumuliert wird. Die Informationsakkumulation ist nicht nur stimulusgeleitet, sondern auch zufälligen Fluktuationen (= noise) unterworfen (Van Zandt, Colonius & Proctor, 2014). Auch wenn kein Stimulus vorhanden ist, kann eine Reaktion ausgelöst werden. Die Information wird für zwei Barrieren gesammelt, eine für Reaktion 1 (z.B.: Tastendruck), eine für Reaktion 2 (z.B.: kein Tastendruck). Wenn eine Barriere erreicht wird, erfolgt eine entsprechende Reaktion. Die Rate der Informationsakkumulation und die Höhe der Barrieren sind variabel (Smith & Ratcliff, 2004).

In Wettlaufmodellen erfolgt die Informationsakkumulation beider Barrieren parallel und unabhängig voneinander (Smith & Ratcliff, 2004). Während ein mentaler Prozess Information für Reaktion 1 sammelt, sammelt zeitgleich ein weiterer Prozess Information für Reaktion 2. Der schnellere Prozess löst eine Reaktion aus. Bei einer richtigen Reaktion war die Informationsakkumulation stimulusgeleitet; der Stimulus wurde richtig als solcher erkannt und die richtige Barriere wurde zuerst erreicht. Bei einer Fehlreaktion war die Informationsakkumulation uninformiert, oder fehlgeleitet (z.B.: Raten, Stimulus nicht erkannt) und die falsche Barriere wurde zuerst erreicht. Findet eine Fehlreaktion statt, ist nicht klar, wie lange die Reaktionszeit für eine richtige Reaktion gewesen wäre. Die Reaktionszeit ist zensiert. Allerdings ist bekannt, dass die richtige Reaktion nach der Fehlreaktion hätte stattfinden müssen.

Reaktion 1 und Reaktion 2 sind wie Loss und Death in der KM unabhängige, parallele Prozesse. Eine Fehlreaktion bzw. ein Loss zensiert in beiden Fällen das interessierende Ereignis. Wird die KM in Reaktionszeitexperimenten angewandt, entspricht  $x$  einer richtigen Reaktion (Tastendruck auf einen Zielreiz) und Loss einer Fehlreaktion (z.B.: Tastendruck auf Stimulus, der keinen Zielreiz darstellen). Fehlreaktionen werden bis zur nächsten langsameren richtigen Reaktionszeit als vorhanden gehandhabt. Danach wird die Wahrscheinlichkeit zu reagieren gleichmäßig auf die nachfolgenden Intervalle redistribuiert.

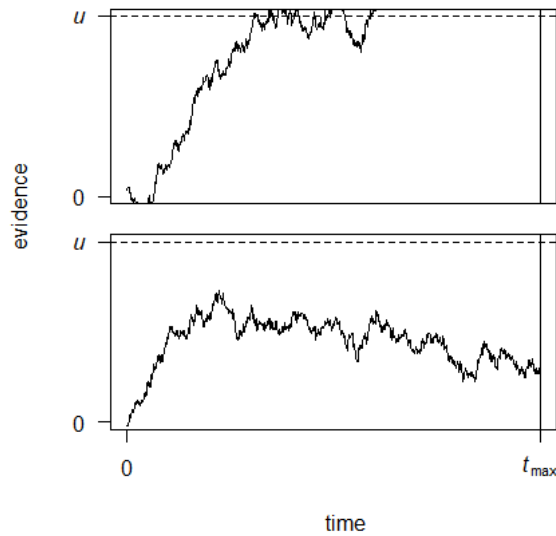


Abbildung 2 beispielhafte Darstellung eines Wettlaufmodells. Der obere Prozess sammelt Information für Reaktion 1, der untere Prozess für Reaktion 2. Die Informationsakkumulation für Reaktion 1 erreicht zuerst die Barriere und würde eine entsprechende Reaktion auslösen.

Die KM-Methode hat im Vergleich zum Reduced-Sample Schätzer einen geringeren Informationsverlust. Während ein Loss in der KM erhalten bleibt, werden bei einem RS-Schätzer die Loss zu Beginn gestrichen (Kaplan & Meier, 1958).

Wird die KM als Fehlerkorrektur auf den fiktiven Datensatz aus Tab. 1 angewandt, wird die mittlere Reaktionszeit von Testperson 2 langsamer als bei der Anwendung des RS (siehe Tab. 3). Die langsamere mittlere Reaktionszeit entspricht der logischen Überlegung, dass Testperson 2 nur nach 280 ms nochmals hätte reagieren können.

Tabelle 3 Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit \*).

Trial	Testperson 1	Testperson 2	Korrektur RS	Korrektur KM
1	200	200	200	200
2	210	210	210	210
3	220	220	220	220
4	230	230	230	230
5	240	240	240	240
6	250	250	250	250
7	260	260	260	260
8	270	270	270	270
9	280	280*	<del>280</del>	295
10	290	290	290	290
11	300	300	300	300
	$\bar{x} = 250$	$\bar{x} = ?$	$\bar{x} = 247$	$\bar{x} = 251,36$

## 4 Aalen-Johansen Schätzer (AJ)

Der AJ entstand wie die Kaplan-Meier Methode (KM) im Zuge von Lebenszeitanalysen. Er verallgemeinert die KM für die Inzidenzfunktion zweier oder mehrerer Klassen von Ereignissen, sogenannte *competing risks* (Johansen, 1978). Die verschiedenen risks schließen sich dabei aus. Wenn beispielsweise ein/e PatientIn in eine andere Stadt zieht, bleibt ungewiss ob und wann sie an der untersuchten Krankheit versterben würde. Die Inzidenzfunktion wird für die unterschiedlichen Ereignisse separat geschätzt und die Zensierungen werden über die Ereignisse gleichermaßen nach rechts verteilt.

In einer Follow-Up Studie werden 10 Personen beobachtet. 6 Personen versterben mit  $t = 10, 20, 30, 40, 50, 60$  Tagen. Eine siebte Person scheidet nach  $t = 70$  Tagen aus der Untersuchung aus (z.B. wegen Umzug in eine andere Stadt). Personen 8 und 10 versterben mit  $t = 80$  und  $90$  Tagen und Person 9 wird mit 100 Tagen gesund entlassen. Über die siebte Person lässt sich nicht mit Sicherheit sagen, wann sie verstorben wäre oder ob sie gesundet wäre. Der AJ verteilt die Wahrscheinlichkeit auf beide Ereignisse inkl. Zeiten, d.h. zu zwei Teilen auf Tod und zu einem Teil auf gesunde Entlassung. Beträgt die Wahrscheinlichkeit zu Versterben vor der Zensierung noch je 10% pro Intervall, beträgt die Wahrscheinlichkeit nach der Zensierung je 13% pro Intervall. Zudem beträgt die Wahrscheinlichkeit zu gesunden ebenfalls 13%.

### 4.1 Implementierung des AJ in Reaktionszeitexperimenten

Diffusionsmodelle haben ebenso wie Wettlaufmodelle zwei Barrieren für die Information akkumuliert wird. Auch die Annahmen über die Informationsakkumulation sind ident (siehe Kap. 3.1). Im Unterschied zu Wettlaufmodellen nehmen Diffusionsmodelle jedoch nur einen mentalen Prozess der Informationsakkumulation an (Abb. 3). D.h. wenn für eine Barriere Information gesammelt wird, wird keine Information für die andere Barriere gesammelt. Die beiden Barrieren und die damit verbundenen Reaktionen liegen somit auf den gegenüberliegenden Enden der Evidenzdimension (Ratcliff, 1978); sie sind negativ korreliert (Smith & Ratcliff, 2004). Findet eine Fehlreaktion nun statt, müsste die Testperson mit der Informationsakkumulation von Null anfangen. Sie würde daher sehr viel, oder unendlich viel Zeit benötigen um noch richtig zu reagieren.

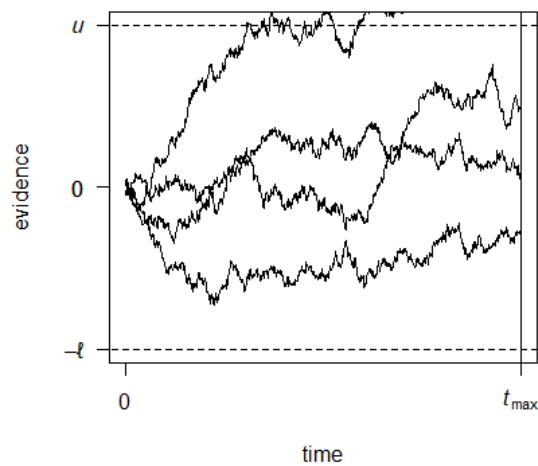


Abbildung 3 beispielhafte Abbildung eines Diffusionsmodelles mit vier mentalen Prozessen. Ein mentaler Prozess akkumuliert Information für eine informierte Reaktion (obere Barriere), die anderen Prozesse erreichen keine Barriere im Beobachtungszeitraum und führen zu keiner Reaktion.

Wertet man ein Reaktionszeitexperiment mit der KM aus, geht man davon aus, dass Fehlreaktionen die Folge zufälliger Störungen sind und ist daran interessiert, was gewesen wäre, wenn es keine Störungen gegeben hätte. Wertet man ein Reaktionszeitexperiment nach dem Diffusionsmodell aus, lässt man die Fehlreaktionen in den Daten und codiert sie z.B. in einem zweiten Auswertungsschritt mit einer besonders langsamen Reaktionszeit, um die „schlechte Leistung“ adäquat abzubilden. Dies entspricht dem Spezialfall des AJ (ohne Zensierungen), bei dem die Ereigniszeit für eines der konkurrierenden Risiken unendlich lang ist. Zensierungen kämen lediglich dann zum Einsatz, wenn man z.B. mit einem variablen Reaktionsfenster arbeiten würde (van Zandt, Colonius & Proctor, 2002).

Im fiktiven Datensatz aus Tabelle 1 hat Testperson 2 eine Fehlreaktion bei 280 ms. Das fiktive Beobachtungsintervall liegt bei 500 ms, daher werden 500 ms anstelle der 280 ms eingesetzt. Daraus ergibt sich eine neue mittlere Reaktionszeit von 270 ms, welche nun auch langsamer ist als die mittlere Reaktionszeit der KM (siehe Tabelle 4). Der AJ kann als strengere Korrektur aufgefasst werden. Beim AJ wird angenommen, dass nach einer Fehlreaktion Information über den Stimulus wieder von 0 weg gesammelt werden müsste. Bei der KM sind informierte Reaktionen und Fehlreaktionen parallel ablaufende Prozesse, wodurch selbst bei einer Fehlreaktion es danach recht schnell zu einer informierten Reaktion kommen kann.

Tabelle 4 *Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit \*).*

TP 2	Korrektur RS	Korrektur KM	Korrektur AJ
200	200	200	200
210	210	260	210
220	220	220	220
230	230	230	230
240	240	240	240
250	250	250	250
260	260	260	260
270	270	270	270
280*	<del>280</del>	280	500
290	290	290	290
300	300	300	300
$\bar{x} = ?$	$\bar{x} = 247$	$\bar{x} = 251,36$	$\bar{x} = 270$

## 4.2 Kritik an sequentiellen-sampling Modellen

Obwohl sequentielle-sampling Modelle Reaktionszeitverteilungen und deren Genauigkeiten gut vorhersagen können, sind sie nicht frei von Schwächen. Erstens sind die Annahmen über Verteilung und Wachstumsrate kaum durch experimentelle Konditionen belegt. Die Annahme einer Gauß'schen Verteilung, wie auch einer Invarianz der Wachstumsraten ist weder empirisch noch theoretisch begründet (Jones & Dzhafarov, 2014). Zweitens können sequentielle-sampling Modelle ineinander übersetzt werden. Es sind nicht die theoretischen, sondern rein die technischen Annahmen die den Vorhersagewert des Modelles bestimmen. Eine Falsifizierbarkeit der einzelnen Modelle sei daher auch nicht möglich (Jones & Dzhafarov, 2014).

## 5 Kill-the-Twin Prozedur (KTT)

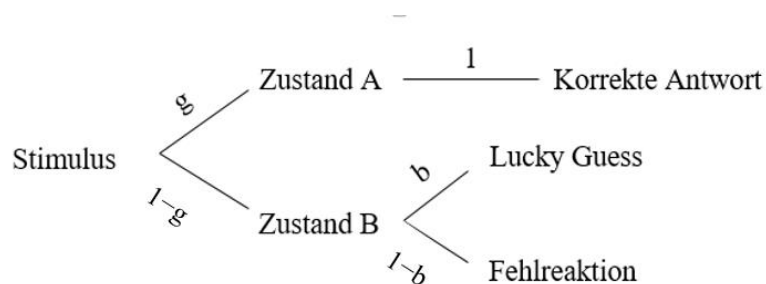
Bei einer Multiple-Choice Prüfung können Studierende entweder die richtige Antwort wissen, oder sie können raten. Wird die Antwort gewusst, wird die richtige Antwort im Antwortbogen angekreuzt. Raten sie jedoch, kann die gegebene Antwort sowohl falsch als auch richtig sein. Ob sie zufälligerweise die richtige Antwort erraten, hängt von der a priori Wahrscheinlichkeit eine Antwort erraten zu können ab. Diese a priori Wahrscheinlichkeit ist von der Anzahl an Antwortmöglichkeiten abhängig. Bei zwei Antwortmöglichkeiten haben Studierende eine Chance von 50% zufälligerweise die richtige Antwort zu wählen. Weiß eine Studentin 80% aller Fragen, bleiben 20% an Fragen in denen sie raten muss. Bei zwei Antwortmöglichkeiten wird sie ca. weitere 10% richtig beantworten, wodurch sie insgesamt ca. 90% der Fragen richtig beantwortet. Hat ein Student 80% aller Fragen in einem Test richtig beantwortet, hat er in 20% falsch geraten. Da er bei zwei Antwortoptionen mit

50% Wahrscheinlichkeit richtig rät, kommen auf 20% falsch geratenen Durchgänge nochmals 20% richtig geratene Durchgänge. Die „wahre“ Performanz liegt daher bei ungefähr 60%. Bei Prüfungen wird zumeist für die Ratewahrscheinlichkeit korrigiert. Obwohl Testpersonen in Reaktionszeitexperimenten auch raten können, stellt eine Ratekorrektur keine Selbstverständlichkeit dar. Die Kill-the-Twin Prozedur (KTT) (Eriksen, 1988) stellt eine mögliche Ratekorrektur für Reaktionszeitexperimente dar.

Die KTT ist eine Ratekorrektur, welche von Eriksen (1988) im Rahmen der Wettlaufungleichung (siehe Kap. 5.1.1 Exkurs: die Wettlaufungleichung) zur Korrektur von Reaktionszeitdaten vorgeschlagen wurde. Fehlreaktionen entsprechen mit der KTT falsch geratenen Durchgängen. In einem 2AFC-Experiment stellen Fehlreaktionen nur 50% der geratenen Durchgänge dar. Die restlichen 50% sind richtig geratene Durchgänge, sogenannte *lucky guesses*, und sind von den anderen richtigen Durchgängen nicht unterscheidbar. Die Verteilung und Anzahl der lucky guesses wird mit der KTT anhand der Fehlreaktionen geschätzt (Gondan & Heckel, 2008).

Der RS nimmt eine Normalverteilung der Ratedurchgänge an. Wären Ratedurchgänge normalverteilt, käme es zu (fast) keiner Verzerrung der mittleren Reaktionszeiten. Evidenz zeigt jedoch, dass die Verteilung rechtsschief ist; die meisten Fehler finden in den schnelleren Durchgängen statt (Van der Heijden et al., 1984; Eriksen, 1966). Somit finden sich gehäuft lucky guesses in den schnelleren Durchgängen. Bleiben diese lucky guesses in den Daten, wird die „wahre“ Performanz einer Testperson überschätzt. Link kritisiert den RS als Schätzer ausdrücklich. Das bloße Reduzieren auf richtige Durchgänge in der Analyse sind „unacceptable capitalizations on chance“ (Link, 1981, S. 485). Die KTT sollte im Vergleich zum RS auch bei einer geringen Fehlerrate validere Ergebnisse liefern.

Stimulusgeleitete Reaktionen, oder uninformierte Reaktionen (z.B.: Raten) können als zwei distinkte mentale Zustände angenommen werden (Ollman, 1966; Yellott, 1971). Zustand A (stimulusgeleitete Reaktion) tritt mit der Wahrscheinlichkeit  $g$  ein. Zustand B (uninformierte Reaktion) tritt mit der Wahrscheinlichkeit  $1-g$  ein. In Zustand A wartet eine Testperson mit ihrer Reaktion bis sie den Stimulus identifiziert hat. Ihre Reaktion sollte richtig sein (Ollman, 1966 zitiert nach Yellott, 1971). In Zustand B wartet die Testperson nicht bis sie den Stimulus identifiziert hat; sie reagiert uninformiert. Die Wahrscheinlichkeit für einen lucky guess ist das Verhältnis von Zielreiz- zu Distraktordurchgängen (Eriksen, 1966) (siehe Abb. 4).



*Abbildung 4* Schematische Darstellung zwei distinkter mentaler Prozesse. Ausgehend von einem Stimulus gelangt Testperson S mit einer Wahrscheinlichkeit von  $g$  in Zustand A, oder mit einer Wahrscheinlichkeit von  $1-g$  in Zustand B. In Zustand A sollte eine 100% richtige Reaktion erfolgen. In Zustand B kann sie mit der Wahrscheinlichkeit  $b$  eine richtige Reaktion treffen, oder mit  $1-b$  eine Fehlreaktion machen.

Um einen lucky guess zu identifizieren gibt es drei Möglichkeiten.

- Erstens kann eine gleichschnelle richtige Reaktionszeit eliminiert werden.
- Zweitens, sofern keine gleichschnelle richtige Reaktion vorhanden ist, kann die nächstlangsamere, richtige Reaktionszeit eliminiert werden.
- Analog hierzu, kann drittens die nächstschnellere, richtige Reaktionszeit eliminiert werden.

Ein rechts-kill (nächstlangsamere, richtige Reaktion wird gekillt) ist konservativer, während ein links-kill (nächstschnellere, richtige Reaktion wird gekillt) ein aggressiveres Vorgehen ist. Ein links-kill ist dahingehend aggressiver, da einer Testperson nicht nur richtige Durchgänge, sondern schnelle richtige Durchgänge eliminiert werden. Ihre mittlere Reaktionszeit wird langsamer als bei einem rechts-kill.

In der vorliegenden Arbeit wird untersucht, ob die Anwendung einer Ratekorrektur ohne (RS) und mit einer Fehlerkorrektur (KM oder AJ) die Reliabilität in Reaktionszeitexperimenten verbessert. Ohne Fehlerkorrektur werden lucky guesses unter Anwendung der KTT aus der Reaktionszeitverteilung gestrichen. Die beobachteten Fehlreaktionen werden als falsch geratene Ratedurchgänge aufgefasst. Mit der KM und dem AJ werden lucky guesses zensiert. Wie die erwarteten Auswirkungen aussehen können, kann mit den Beispieldaten von Tabelle 1 veranschaulicht werden (siehe Tab. 5).



Tabelle 5 Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit \*).

TP 2	RS	KM	AJ	KTT <sub>r</sub>	KM <sub>KTT</sub>	AJ <sub>KTT</sub>
200	200	200	200	200	200	200
210	<del>210</del>	210	210	210	210	210
220	220	220	220	220	220	220
230	230	230	230	230	230	230
240	240	240	240	240	240	240
250	250	250	250	250	250	250
260	260	260	260	260	260	260
270	270	270	270	270	270	270
280*	<del>280</del>	295	500	<del>280</del>	300	500
290	290	290	290	<del>290</del>	300	500
300	300	300	300	300	300	300
$\bar{x} = ?$	$\bar{x} = 247$	$\bar{x} = 251,35$	$\bar{x} = 270$	$\bar{x} = 242,22$	$\bar{x} = 252,72$	$\bar{x} = 289,09$

### 5.1 Exkurs: die Wettlaufungleichung.

Die Wettlaufungleichung von Miller (1982) wurde zur Erklärung des Redundanzeffekts formuliert. Als Redundanzeffekt wird das Phänomen bezeichnet, dass Reaktionen auf redundante Signale (z.B.: akustischer und visueller Reiz) schneller erfolgen, als auf einzelne Signale (z.B.: akustischer oder visueller Reiz). Zwei Modelle können diesen Effekt erklären: das Wettlaufmodell oder das Koaktivierungsmodell.

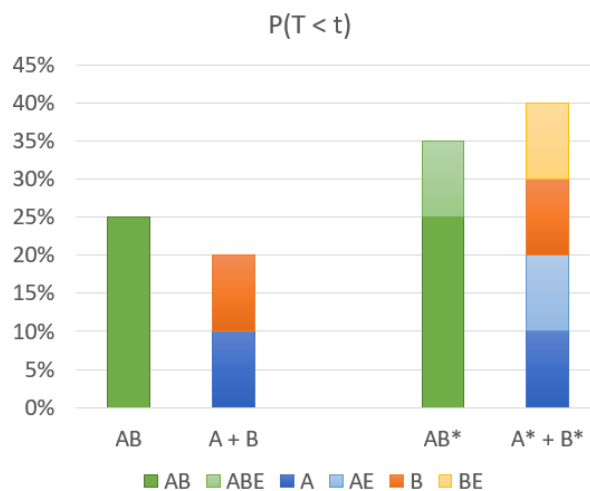
Das Wettlaufmodell postuliert eine separate Verarbeitung von Signalen. Redundante Signale produzieren voneinander unabhängige Aktivierungen und die Reaktion erfolgt durch die schnellere Aktivierung (Eriksen, 1988). Die Wahrscheinlichkeit bei Durchgängen mit redundanten Signalen unter einem kleinen t zu reagieren, kann nicht größer sein als die Summe der Wahrscheinlichkeiten von Durchgängen mit nur einem Zielreiz unter t zu reagieren (Miller, 1982).

$$P(T \leq t | S_1 \text{ and } S_2) \leq P(T \leq t | S_1) + P(T \leq t | S_2) \quad (1)$$

Das Koaktivierungsmodell postuliert stattdessen eine gemeinsame Aktivierung bei redundanten Signalen. Die Ungleichung kann verletzt sein.

Die Wettlaufungleichung ist zugunsten des Wettlaufmodells verzerrt (Eriksen, 1988). Ratedurchgänge (Fehlreaktion oder lucky guess), die vermehrt bei schnelleren Reaktionen stattfinden, bleiben unberücksichtigt. Durch schnelle lucky guesses erhöht sich die Wahrscheinlichkeit unter t zu reagieren für redundante Durchgänge sowie für jeden Durchgang mit nur einem Zielreiz. Der rechte Term profitiert doppelt von lucky guesses, wodurch es schwerer zu einer Verletzung der Ungleichung kommt. Die KTT wird von Eriksen (1988) als Ratekorrektur vorgeschlagen um diesem Ungleichgewicht entgegenzuwirken.

25% aller nicht-geratenen Reaktionen bei redundanten Signalen (AB) finden unter  $t = 300$  ms statt. In Durchgängen mit nur einem Signal (A oder B) finden je 10% aller nicht-geratenen Reaktionen unter  $t = 300$  ms statt. Insgesamt sind nur 20% der A oder B-Durchgänge schneller als das gewählte  $t$ . Die Wettlaufungleichung wäre in diesem Fall verletzt und ein Koaktivierungsmodell müsste angenommen werden. Die Wettlaufungleichung ist nicht mehr verletzt, wenn auf jeden Durchgangstyp (AB, A, B) 10% an lucky guesses unter  $t = 300$  ms kommen. In AB-Durchgängen beträgt die Wahrscheinlichkeit unter  $t = 300$  ms zu reagieren nun 35%, während in A oder B-Durchgängen je 20% aller Reaktionen unter  $t = 300$  ms stattfinden. Insgesamt sind dann 40% der A oder B-Durchgänge schneller als das gewählte  $t$ . Die Wettlaufungleichung wäre bestätigt, ein Wettlaufmodell würde angenommen werden, obwohl die „wahre“ Performanz ein Koaktivierungsmodell nahelegen würde (Abb. 5).



*Abbildung 5* Graphische Darstellung der Wahrscheinlichkeit unter  $t$  zu reagieren. AB sind redundante Durchgänge, A oder B sind jeweils Durchgänge mit einem Zielreiz. Mit \* markierte Durchgänge inkludieren 10% lucky guesses. Während ohne lucky guesses ein Koaktivierungsmodell gelten würde, müsste aufgrund der lucky guesses ein Wettlaufmodell angenommen werden.

## 6 Annahmen

Mit KM und AJ werden Fehlreaktionen nicht aus den Daten gestrichen, sondern durch theoriegeleitete langsamere Reaktionszeiten ersetzt. Dementsprechend sind die mittleren Reaktionszeiten mit KM und AJ langsamer im Vergleich zu den mittleren Reaktionszeiten des RS.

Sofern die Möglichkeit zum Raten berücksichtigt wird, kann folgende Abstufung der mittleren Reaktionszeiten angenommen werden:  $RS \leq KM \leq (AJ \text{ oder } KM_{KTT}) \leq AJ_{KTT}$  (Abb. 6).

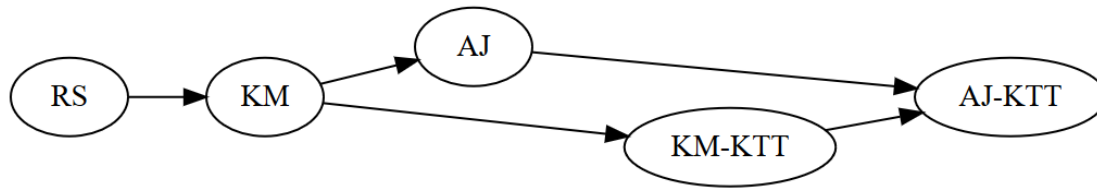


Abbildung 6 Angenommene Abfolge der mittleren Reaktionszeiten

Bei einer Reduktion auf richtige Reaktionen würde Testperson 2 mit einer Fehlreaktion bei 280 ms schneller reagieren, als Testperson 1, die nur richtig reagiert hat. Würde Testperson 2 nochmals reagieren können, könnte sie dies nur noch nach 280 ms machen. Eine kürzere mittlere Reaktionszeit im Vergleich zu Testperson 1 steht im Widerspruch zu logischen Überlegungen. Fehlreaktionen brauchen eine Korrektur und können nicht einfach ausgelassen werden. Bei der KM sind Fehlreaktionen und informierte Reaktionen unabhängig voneinander, beim AJ konkurrieren beide Prozesse miteinander. Unter Hinzunahme der KTT kann zusätzlich um Ratedurchgänge kontrolliert werden.

KM und AJ sollten auch bei einer Retestreliaibilität, oder Split-Half-Reliaibilität wegen der höheren Varianz der „true scores“ besser abschneiden. Unter Inklusion von Ratedurchgängen kann folgende Abstufung für die Reliaibilität angenommen werden:  $RS \leq KM \leq KM_{KTT}$  und  $RS \leq AJ \leq AJ_{KTT}$ . Welche der beiden Fehlerkorrekturen mit und ohne Ratekorrektur reliablere Ergebnisse liefert, ist nicht vorhersehbar, da unbekannt ist, ob Fehlreaktionen und informierte Reaktionen unabhängig voneinander oder korreliert sind. Des Weiteren ist zu beachten, dass AJ und KM nur dann reliablere Ergebnisse liefern können, wenn die Ergebnisse von Testpersonen eine geringe Genauigkeit aufweisen. Bei einer hohen Genauigkeit unterscheiden sich AJ und KM mathematisch kaum vom RS. Es sind daher ähnliche mittlere Reaktionszeiten, sowie ähnliche Reliaibilitäten erwartbar.

Für die vorliegende Arbeit wurde die mediane Reaktionszeit verwendet. Durch Anwendung des AJ werden Fehlreaktionen unendlich gesetzt, wodurch eine mittlere Reaktionszeit nicht mehr berechnet werden kann. Beziehungsweise dauert eine richtige Reaktion nach einer Fehlreaktion laut Diffusionsmodell unendlich lang, wodurch eine Ordinalskala und keine Intervallskala vorliegt. Für die Reliaibilitätsberechnungen wurde die Spearman-Rangkorrelation genommen.

## 7 Methode und Material

Um die in Kapitel 6 formulierten Annahmen zu überprüfen, wurden Retestreliaibilität und Split-Half-Reliaibilität bei vier Untertests der Testbatterie zur Aufmerksamkeitsprüfung (TAP; Zimmermann & Fimm, 2002) mit den oben beschriebenen Schätzverfahren berechnet. Die TAP ist ein neuropsychologisches Instrument für die Diagnostik unterschiedlicher Aufmerksamkeitsfunktionen (u.a. Wachsamkeit, Flexibilität, geteilte Aufmerksamkeit). Die Firma PSYTEST (Psychologische Testsysteme, Herzogenrath) stellte freundlicherweise vier Datensätze zur Verfügung, bei denen mehr als 10% Fehler gemacht wurden: *visual search*, *divided attention*, *go-nogo* und *vigilance*. Die Datensätze des visual search task ( $N = 90$ ;  $> 70$  Jahre) sowie der divided attention ( $N = 185$ , 40–50 Jahre) stammen aus einer Fahrtrainingsstudie. Keine der Testpersonen berichtete relevante neurologische Erkrankungen. Es gab zwei Testzeitpunkte (T1, T2) mit einem Jahr Abstand. Der go-nogo- und der vigilance-Datensatz stammen aus einer Kindernormstichprobe ( $N = 375$ , Alter unklar).

### 7.1 visual search task

Im visual search task müssen die Testpersonen einen Zielreiz in einer 5 x 5 Matrix suchen und darauf per Tastendruck reagieren. Der Zielreiz ist ein nach oben hin geöffnetes Quadrat. Die Distraktoren sind ebenfalls Quadrate, welche allerdings nach links, rechts, oder unten hin geöffnet sind (Abb. 7).

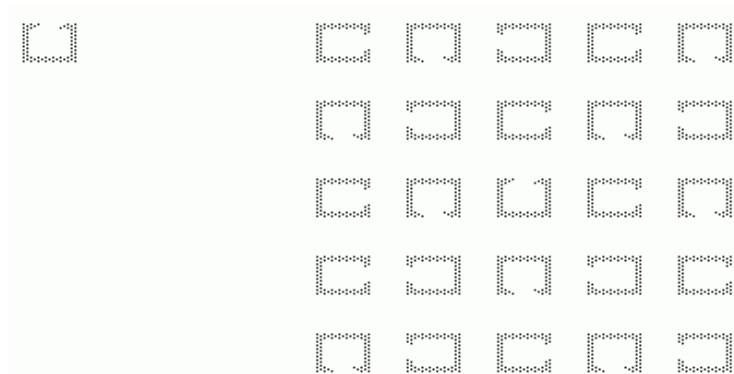


Abbildung 7 Aufgabenbildschirm des visual search task. In dieser 5 x 5 Matrix muss nach einem Quadrat, welches nach oben hin geöffnet ist (hier in der Mitte) gesucht werden. Die Distraktoren bestehen ebenso aus geöffneten Quadraten.

In Zielreiz-Durchgängen ist ein Zielreiz vorhanden und es muss mit Taste 1 reagiert werden. Die Zielreize treten pro Position zwei Mal auf. In Distraktor-Durchgängen sind nur Distraktoren vorhanden und es muss mit Taste 2 reagiert werden. Zielreiz-, wie auch Distraktor-Durchgänge

kommen 50-mal vor. Die Testpersonen werden dazu aufgefordert die Zielreize von links oben nach rechts unten zu suchen, einmal zeilenweise, einmal spaltenweise (Zimmermann & Fimm, 2016).

### 7.1.1 Implementierung der KTT im visual search task.

In Reaktionszeitexperimenten mit nur einer Bedingung eliminiert/zensiert die Kill-The-Twin Prozedur (KTT) für jede Fehlreaktion eine ähnliche, richtige Reaktion in dieser Bedingung. Sind zwei Versuchsbedingungen vorhanden, wie im visual search task, stellt sich die Frage: Wenn eine Testperson in Bedingung 1 falsch reagiert, sind die lucky guesses dann in Bedingung 1 oder 2 zu finden (Tab. 6)?

Tabelle 6. Vierfelderschema der Ratedurchgänge im visual search task.

Versuchsbedingung	Geratene <i>target response</i> (Taste 1)	Geratene <i>non-target response</i> (Taste 2)
Zielreiz	lucky guess	Fehlreaktion
Distraktor	Fehlreaktion	lucky guess

Im visual search task wird die KTT nicht innerhalb einer Bedingung, sondern innerhalb eines Reaktionstyps (Tastendruck) durchgeführt. Die KTT in Tabelle 6 erfolgt spaltenweise und nicht zeilenweise.

Reagiert jemand spontan mit Taste 1 in 10% der Trials, führt dies in 10% der Distraktor-Durchgänge zu einer falsch-positiven Reaktion und in 10% der Zielreiz-Durchgänge zu einem lucky guess. Der Twin einer Fehlreaktion in der Distraktor-Bedingung befindet sich demnach nicht in der Distraktor-Bedingung, sondern in der Zielreiz-Bedingung.

Reagiert jemand spontan mit Taste 2 in 15% der Trials, so führt auch dies wieder in 15% der Zielreiz-Durchgänge zu einer falsch-negativen Reaktion und in 15% der Distraktor-Durchgänge zu einem lucky guess. Der Twin der Fehlreaktion in der Zielreiz-Bedingung befindet sich in der Distraktor-Bedingung.

## 7.2 divided attention

Der Untertest zur divided attention liegt in zwei Testformen vor, die sich an ihrer visuellen Suchaufgabe unterscheiden. Sowohl Testform 1 als auch Testform 2 beinhalten zwei simultane Suchaufgaben, eine visuelle, eine akustische Suchaufgabe. In beiden Testformen muss in der

akustischen Suchaufgabe auf Unregelmäßigkeiten in der Tonsequenz per Tastendruck reagiert werden (Zimmermann & Fimm, 2016).

In Testform 1 muss auf das Auftauchen von 6-8 Kreuze reagiert werden. Genau dann, wenn vier dieser Kreuze die Ecken eines Quadrates bilden, muss reagiert werden. Die Kreuze können an jeder Stelle der 4 x 4 Matrix auftreten (Zimmermann & Fimm, 2016) (Abb. 8).

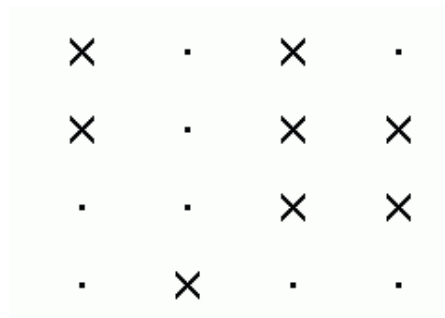


Abbildung 8 Aufgabenbildschirm der divided attention der visuellen Zielreize. Die 6-8 Kreuze können zufällig an jeder Stelle der Matrix auftauchen. Bilden sie, wie in diesem Fall, ein Quadrat, so muss per Tastendruck reagiert werden.

In Testform 2 muss auf 01 und 10 reagiert werden. Die Stimuli werden in der Bildschirmmitte präsentiert und tauchen zu festgelegtem Rhythmus auf. Zwischen den Stimuli können auch um 90° rotierte S als Distraction auftreten. Wird ein S präsentiert soll nicht reagiert werden (Abb. 9).



Abbildung 9 Alternativer Aufgabenbildschirm der divided attention. Die vier Abbildungen zeigen die vier möglichen Stimuli. Zielreize sind hierbei lediglich 01 und 10. Jeder Stimuli taucht in der Mitte des Bildschirmes auf.

Aus dem Untertest zur divided attention kann den Teilnehmenden auch jeweils nur eine der beiden Suchaufgaben vorgelegt werden (Zimmermann & Fimm, 2016). Die zur Verfügung gestellten Datensätze zur divided attention stammen aus der visuellen Suchaufgabe der Testform 1.

### 7.2.1 Implementierung der KTT in divided attention.

Ähnlich zum visual search task wird die KTT nicht innerhalb einer Bedingung, sondern innerhalb eines Reaktionstyps durchgeführt. Typ 1 ist die Reaktion per Tastendruck auf einen Zielreiz. Typ 2 ist keine Reaktion auf keinen Zielreiz. Reagiert jemand spontan mit der festgelegten Taste in 5% der

Trials, führt dies in 5% der Zielreiz-Durchgänge zu einem lucky guess und in 5% der Distraktor-Durchgänge zu einer falsch-positiven Reaktion. Der Twin einer Fehlreaktion in der Distraktor-Bedingung befindet sich demnach in der Zielreiz-Bedingung.

Rät jemand in 10% der Fälle, dass kein Target vorhanden war, führt dies zu 10% Auslassungen in den Zielreiz-Durchgängen und zu 10% lucky guesses in den Distraktor-Durchgängen. Da in den Distraktor-Durchgängen sowieso keine Reaktionszeiten erhoben wurden, wird keine KTT durchgeführt.

Die KTT muss auf die Verteilung von Zielreiz- zu Distraktordurchgängen der divided attention angepasst werden. Der Untertest besteht aus 100 Durchgängen, davon 17 Zielreize und 83 Distraktoren. Daraus ergibt sich ein Verhältnis der Durchgänge von ca. 5 : 1. Eine 1 : 1 Ratekorrektur, wie sie mit der KTT der Fall wäre, würde die Anzahl an lucky guesses überschätzen. In der Anwendung werden hierfür je 5 nach der Größe nach sortierten Fehlreaktionen gruppiert und zu einem Intervall zusammengefasst. Ein lucky guess kann innerhalb, oder außerhalb des Intervalls liegen. Es kann eine aggressive, oder konservative KTT angewandt werden.

### 7.3 go-nogo

Im Untertest go-nogo müssen Testpersonen zugleich auf Zielreize reagieren, als auch Distraktoren ignorieren. Nur auf Zielreize soll ein Tastendruck erfolgen. Es gibt zwei mögliche Testformen.

In Testform 1 gibt es zwei Stimuli ( $x$  und  $+$ ) von denen  $x$  den Zielreiz darstellt. Beide Stimuli tauchen in der Bildschirmmitte zu einem festgelegten Intervall auf und beide Stimuli kommen je 20-mal vor (Abb. 10). Die Stimuli sind leicht auseinanderzuhalten und sollen eine impulsive Reaktion evozieren (Zimmermann & Fimm, 2016). Die zur Verfügung gestellten Datensätze liegen in Testform 1 vor.



*Abbildung 10* Stimuli des Subtest go-nogo in der ersten Bedingung. Bei einem  $x$  muss per Tastendruck reagiert werden. Bei einem  $+$  soll nicht reagiert werden.

In Testform 2 gibt es fünf Stimuli (Quadrate mit unterscheidbaren Mustern). Zwei Quadrate werden vorab als Zielreiz festgelegt. Auf diese Zielreize soll abermals mit Tastendruck reagiert werden, auf die anderen drei nicht. Die Stimuli werden nacheinander in der Bildschirmmitte präsentiert, Zielreize 24-mal, Distraktoren 36-mal (Abb. 11) (Zimmermann & Fimm, 2016).

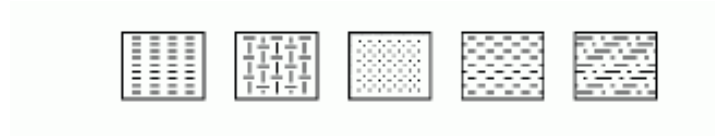


Abbildung 11 Die 5 Stimuli aus dem Subtest go-nogo. 2 von ihnen werden vorab als mögliche Zielreize definiert. Nur auf diese soll per Tastendruck reagiert werden.

Die KTT wird analog zur KTT in divided attention verwendet, allerdings mit einem 1 : 1 kill.

## 7.4 vigilance

Es gibt drei mögliche Testformen für vigilance: eine rein auditorische Bedingung, eine visuelle Bedingung namens *jumping square* und eine weitere visuelle Bedingung namens *moving bar* (Zimmermann & Fimm, 2016). Die kritischen Stimuli werden in einer niedrigen Frequenz unter eine monotone Aufgabe dargeboten.

In der auditorischen Bedingung wird eine irreguläre Sequenz von hohen und niedrigen Tönen dargeboten. Eine Testperson muss per Tastendruck reagieren, wenn entweder ein hoher Ton auf einen hohen Ton, oder ein niedriger Ton auf einen niedrigen Ton folgt. Testpersonen müssen dann reagieren, wenn die Irregularität der Sequenz unterbrochen ist.

In der visuellen Bedingung *jumping square* besteht das Ziel abermals in der Detektion der Änderung einer Sequenz. In der Mitte des Bildschirmes werden zwei übereinanderstehende Quadrate präsentiert. In beiden Quadraten kann ein Muster erscheinen. Scheint das Muster zweimal im selben Quadrat auf, muss die Testperson per Tastendruck reagieren.

In der visuellen Bedingung *moving bar* ist der Stimuli ein sich vertikal bewegendes Balken. Die vertikale Bewegung ausgehend von der Bildschirmmitte kann dabei variieren. Ist die Auslenkung nach unten, oder oben hin größer als in die andere Richtung muss per Tastendruck reagiert werden.

Per Testform gibt es 36 kritische Stimuli von insgesamt 1200 Stimuli in der auditorischen Bedingung und *jumping square*, sowie 2800 Stimuli in *moving bar*. Die verwendete Testform ist *jumping square*.

Die KTT wird analog zur Implementierung in divided attention verwendet. Das Verhältnis von Zielreiz zu Distraktor beträgt allerdings 33:1.



## 7.5 Statistische Auswertung

Die statistische Auswertung der Retestreliaibilität und Split-Half-Reliaibilität erfolgte über R Version 3.4.3 (R Core Team, 2018). Neben dem Standardpackage wurden zudem die Packages *gdata* und *survival* verwendet. Retestreliaibilität und Split-Half-Reliaibilität wurden unter Berücksichtigung aller Fälle über die Spearman-Korrelation berechnet.

Um die mögliche Wirksamkeit von Fehler- und Ratekorrekturen abschätzen zu können, wurde die Genauigkeit zu beiden Messzeitpunkten, beziehungsweise Stichprobenhälften berechnet. Bei hoher Genauigkeit unterscheiden sich die Schätzungen von RS, KM, AJ und KTT kaum. Demnach ist auch kein Unterschied in den Schätzungen der Reliaibilität erwartbar.

### 7.5.1 Implementierung der Schätzer.

Die interessierende Variable ist die mediane Reaktionszeit. Wenn zwei Messzeitpunkte vorhanden sind, wird die Retestreliaibilität berechnet. Bei nur einem Messzeitpunkt wird die Split-Half-Reliaibilität verwendet. Für die Split-Half-Reliaibilität wird pro Testperson eine Zufallsreihenfolge der Durchgänge erstellt (jede Testperson hat eine andere Reihenfolge) und dann werden die geraden mit den ungeraden Durchgängen verglichen. Beide Reliaibilitätsmaße sind Spearman-Rangkorrelationen. Pro Subtest werden 6 Varianten der Reaktionszeitverteilung (und damit auch der medianen Reaktionszeit) berechnet: RS, RS<sub>KTT</sub>, KM, KM<sub>KTT</sub>, AJ, AJ<sub>KTT</sub>. Fehlreaktionen des AJ werden auf unendlich gesetzt. Mediane, die mit der KM nicht berechnet werden können, werden ebenfalls unendlich gesetzt.

Es werden zwei Formen der KTT verwendet, eine aggressive und eine konservative Form. Bei einer aggressiven KTT wird der nächst-schnellere richtige Durchgang identifiziert und ebenfalls eliminiert/zensiert. Zusätzlich erfolgt eine Zensierung des nächst-langsameren richtigen Durchganges, falls der schnellste Durchgang eine Fehlreaktion war. Bei einer konservativen KTT wird der nächst-langsamere richtige Durchgang identifiziert und ebenfalls eliminiert/zensiert.

## 8 Ergebnisse

### 8.1 Retestreliaibilität des visual search task

Es lagen 90 anonymisierte Datensätze mit Messwiederholung (T1, T2) von SeniorInnen aus einer Fahrtrainingsstudie vor. Die Daten wurden freundlicherweise von der Firma PSYTEST (Psychologische Testsysteme, Herzogenrath) zur Verfügung gestellt. Das Alter der Testpersonen lag

über 70 Jahren, weitere ProbandInnencharakteristika lagen nicht vor. Da sich die Aufgaben in den beiden Testbedingungen (*Zielreiz*, *Distraktor*) unterscheiden und sich dementsprechend auch die Informationsverarbeitung unterscheidet, wurde die Retestreliaibilität für beide Versuchsbedingungen separat berechnet (Tabelle 7).

Tabelle 7. Retestreliaibilität per RS, KM,  $KM_{KTT}$ , AJ und  $AJ_{KTT}$  sowie die Genauigkeit zu T1, T2.

N = 90	Accuracy		RS			KM			AJ		
	T1	T2	keine	kons*	agg	keine	kons	agg	keine	kons	agg
<i>Zielreiz</i>	85%	87%	.68	.64	.68	.73	.73	.73	.74	.75	.75
<i>Distraktor</i>	87%	90%	.73	.72	.72	.73	.76	.77	.73	.75	.75

\*keine: keine KTT, kons: konservative KTT, agg: aggressive KTT

Die Schätzungen für die Retestreliaibilität unterscheiden sich ein wenig in Abhängigkeit vom verwendeten Schätzer und der Testbedingung. In den Zielreiz-Durchgängen ist die Reliaibilität mit Fehlerkorrektur (KM, AJ) generell höher als der RS-Schätzer. In Kombination mit der Ratekorrektur (KTT, aggressiv oder konservativ) kommt es wiederum zu einer minimalen Verbesserung. Eine alleinige Ratekorrektur führt allerdings zu einer Verringerung der Retestreliaibilität. In den Distraktor-Durchgängen unterscheidet sich die Retestreliaibilität nur geringfügig zwischen den verschiedenen Schätzmethoden. Dies ist zum Teil auf die hohe Genauigkeit zu beiden Testzeitpunkten zurückzuführen, sodass die Fehlerkorrekturen alleine nur zu wenig geänderten Ergebnissen führen. Allerdings lässt sich eine minimale Verbesserung der Reliaibilität für die Kombination KM mit KTT und AJ mit KTT beobachten.

## 8.2 Retestreliaibilität der divided attention

Es lagen insgesamt 188 anonymisierte Datensätze mit Messwiederholung (T1, T2) aus derselben Fahrtrainingsstudie vor. Die Daten wurden abermals von der Firma PSYTEST (Psychologische Testsysteme, Herzogenrath) zur Verfügung gestellt. Das Alter der Testpersonen lag zwischen 40 und 50 Jahren, weitere ProbandInnencharakteristika lagen nicht vor. Bei drei der Testpersonen lag kein Datensatz zum zweiten Testzeitpunkt vor. Die Reliaibilität wurde daher mit einem reduzierten Datensatz von 185 berechnet.

Tabelle 8. *Retestreliaibilität per RS, KM, KM<sub>KTT</sub>, AJ und AJ<sub>KTT</sub> sowie die Genauigkeit zu T1, T2.*

N = 185	Accuracy		RS			KM			AJ		
	T1	T2	keine	kons*	agg	keine	kons	agg	keine	kons	agg
	52%	62%	.34	.28	.30	.35	.37	.37	.33	.36	.36

\*keine: keine KTT, kons: konservative KTT, agg: aggressive KTT

Die Retestreliaibilität unterscheidet sich in Abhängigkeit des verwendeten Schätzers kaum. Fehlerkorrekturen führen zu keiner nennenswerten Verbesserung der Reliaibilität. Während die KM tendenziell eine höhere Reliaibilität zeigt, ist die Reliaibilität der AJ tendenziell geringer im Vergleich zum RS. Der Einsatz einer Ratekorrektur verschlechtert die Reliaibilität des RS, führt allerdings zu einer Steigerung der Reliaibilität mit KM und AJ. Die Reliaibilität bei einer Kombination aus Fehler- und Ratekorrektur ist höher als beim RS. Aufgrund der niedrigen Genauigkeit zu T1 (52%) und T2 (62%) wäre erwartbar gewesen, dass Fehler- und Ratekorrekturen eine höhere Reliaibilität ermöglichen. Dem ist allerdings nicht der Fall.

### 8.3 Split-Half-Reliaibilität des go-nogo

Es lagen insgesamt 375 anonymisierte Datensätze einer Normstichprobe des TAP ohne Messwiederholung vor. Anstelle der Retestreliaibilität wurde demnach die Split-Half-Reliaibilität (Benennung: S1, S2) berechnet. Die ProbandInnen waren Kinder (keine genaueren Angaben vorhanden). Bei 12 ProbandInnen waren die Daten des go-nogo nicht vorhanden. 363 Datensätze wurden für die Auswertung verwendet.

Tabelle 9. *Split-Half-Reliaibilität RS, KM, KM<sub>KTT</sub>, AJ und AJ<sub>KTT</sub> sowie die Genauigkeit zu T1, T2.*

N = 363	Accuracy		RS			KM			AJ		
	S1	S2	keine	kons*	agg	keine	kons	agg	keine	kons	agg
	85%	86%	.74	.68	.70	.76	.72	.77	.76	.48	.49

\*keine: keine KTT, kons: konservative KTT, agg: aggressive KTT

Die Split-Half-Reliaibilität unterscheidet sich ein wenig in Abhängigkeit des verwendeten Schätzers. Die Fehlerkorrekturen (KM, AJ) weisen tendenziell eine höhere Reliaibilität als der RS auf. Eine Ratekorrektur verschlechtert die Reliaibilität tendenziell mit jedem Schätzer. Die Kombinationen RS<sub>KTT</sub> und AJ<sub>KTT</sub> weisen gegenüber RS und AJ eine geringere Reliaibilität auf. In Kombination mit der KM verringert sich die Reliaibilität zwar bei einer konservativen KTT, verbessert sich allerdings

minimal bei einer aggressiven KTT. Allerdings verringert die KTT tendenziell auch hier die Reliabilität.

### 8.4 Split-Half-Reliabilität der vigilance

Es lagen 375 anonymisierte Datensätze der Normstichprobe von go-nogo ohne Messwiederholung vor. Zur Berechnung der Reliabilität wurde die Split-Half-Reliabilität angewandt. Die ProbandInnen waren Kinder (keine genaueren Angaben vorhanden). Bei 19 Testpersonen war der Untertest vigilance nicht vorhanden. Es wurde daher mit einer reduzierten Stichprobe von 356 Datensätzen gerechnet.

Tabelle 10. *Split-Half-Reliabilität von RS, KM, KM<sub>KTT</sub>, AJ und AJ<sub>KTT</sub> sowie die Genauigkeiten zu T1 und T2.*

N = 356	Accuracy		RS			KM			AJ		
	S1	S2	keine	kons*	agg	keine	kons	agg	keine	kons	Agg
	83%	83%	.83	.86	.83	.85	.87	.87	.85	.87	.86

\*keine: keine KTT, kons: konservative KTT, agg: aggressive KTT

Die Split-Half-Reliabilität unterscheidet sich wenig in Abhängigkeit des verwendeten Schätzers. Tendenziell weisen Fehlerkorrekturen (KM, AJ) im Vergleich zum RS eine höhere Reliabilität auf. Eine Ratekorrektur führt ebenfalls zu leicht höheren Reliabilitäten.

## 9 Diskussion

Unter Anwendung von Fehler- und Ratekorrekturen verbessert sich die Reliabilität in allen Untertests der TAP geringfügig. KM und AJ schneiden bei der Schätzung von Retestreliabilität und Split-Half-Reliabilität im Vergleich zum RS besser ab. Sie dürften eine genauere Schätzung der Varianz der „true scores“ ermöglichen. Die Ergebnisse für die KTT sind nicht eindeutig. Tendenziell könnten folgende Abstufungen gelten:  $RS \leq KM \leq KM_{KTT}$  und  $RS \leq AJ \leq AJ_{KTT}$ .

Die angenommenen Abstufungen der Reliabilität lassen sich für visual search task, divided attention (mit Uneindeutigkeit bei den Fehlerkorrekturen) und vigilance finden. In go-nogo verbessern KM und AJ zwar die Reliabilität, allerdings verschlechtert die KTT die Reliabilität in Kombination mit jedem Schätzer. Die Verschlechterung der Reliabilität in Kombination RS mit KTT war erwartbar. Mit dieser Kombination werden nicht nur Fehlreaktionen, sondern auch deren angenommene Twins

eliminiert, wodurch sich die Varianz verringert. Verringert sich die Varianz, verringert sich ebenso die Reliabilität.

Bei der KM und der AJ erfolgt zwar ebenfalls eine Informationsreduktion, allerdings bleibt eine höhere Varianz erhalten. Die Reliabilität sollte anders als bei  $RS_{KTT}$  nicht sinken, in go-nogo sinkt die Reliabilität allerdings auch in der Kombination von Fehler- mit Ratekorrektur. Eine mögliche Erklärung könnte sein, dass Testpersonen in go-nogo nicht raten. Eine andere Erklärung kann sein, dass die verwendete Kill-Methode (zeitliche nahe Twins eliminieren/zensieren) in go-nogo keine adäquate Ratekorrektur darstellt.

In divided attention unterscheiden sich die Reliabilitäten hinsichtlich des verwendeten Schätzers nicht, beziehungsweise weisen die Ergebnisse keine klare Tendenz auf. Aufgrund der geringen Genauigkeit sollten Fehlerkorrekturen allerdings reliablere Ergebnisse liefern. Bei einer geringen Genauigkeit wäre mehr Spielraum für KM und AJ vorhanden. In der divided attention gibt es jedoch keinen Unterschied in der Reliabilität zwischen den Schätzern.

Im Unterschied zu den anderen verwendeten Subtests sind die Reaktionszeiten in divided attention weder von großer Bedeutung noch als reliables Maß konzipiert. Sie sind lediglich von „secondary importance“ (Zimmermann & Fimm, 2016, S. 27). Dieser Umstand könnte einen Teil der fehlenden Wirksamkeit von KM und AJ erklären. Wenn die Reaktionszeiten nicht als reliabel konzipiert und für die Auswertung des Testes nicht wichtig sind, dann kann die „wahre“ Performanz von Testpersonen selbst mit Fehlerkorrekturen nicht gut geschätzt werden.

Fehler- und Ratekorrekturen führen in der divided attention zwar zu keiner Steigerung der Reliabilität, eine Dichotomisierung der Daten allerdings schon. Die Dichotomisierung erfolgte in unserem Fall anhand der Genauigkeit. Die Genauigkeit entspricht dem Mittelwert der richtigen Reaktionen über alle Durchgänge. In einer explorativen Analyse zeigte sich eine höhere Reliabilität der Genauigkeit ( $r = 0.38$ ) im Vergleich zu RS, KM und AJ. Die Reliabilität der Genauigkeit ließ sich auch durch die KTT tendenziell erhöhen ( $r = 0.39$ ).

Eine Dichotomisierung von Daten kann vor allem bei vorhandener Kontamination sinnvoll sein (Shentu & Xie, 2010). Üblicherweise wird eine Dichotomisierung von Daten vermieden, da es zu einer Informationsreduktion kommt. Gleichzeitig reduziert sich ebenfalls die Kontamination. Wenn die Reduktion der Kontamination größer ist als die Informationsreduktion, kann eine Dichotomisierung den Verlust an Information aufwiegen (Shentu & Xie, 2010). Dichotomisierte Maße können in dem Fall dann z.B. reliablere Schätzungen ermöglichen. Eine Dichotomisierung ist vor allem dann sinnvoll, wenn kontaminierte Daten vorliegen und diese nicht neu erhoben werden können, oder wenn die Ursache der Verzerrung unbekannt ist. In der divided attention dürfte die

Dichotomisierung etwaige Kontaminierungen mehr reduziert haben als sie Information reduziert hat. Daher ist die Reliabilität der Genauigkeit höher im Vergleich zu RS, KM und AJ.

Die Ergebnisse zeigen zwar die angenommene Abstufung der Reliabilität ( $RS \leq KM \leq KM_{KTT}$  und  $RS \leq AJ \leq AJ_{KTT}$ ), allerdings ist der Unterschied in den Schätzungen der Reliabilität eher klein. Ausgehend hiervon kann sich die Frage gestellt werden, warum dann der Reduced-Sample Schätzer nicht weiterhin verwendet werden sollte.

Auch wenn die gefundenen Unterschiede klein sind, ist die Verwendung des RS dennoch problematisch. Fehlreaktionen werden als uninformativ gehandhabt, was diese allerdings nicht sind. Eine Fehlreaktion zum Zeitpunkt  $t$  bedeutet, dass die Testperson die Antwort noch nicht weiß (vgl. Kaplan-Meier-Methode), oder nie wissen wird (vgl. Diffusionsmodell/Aalen-Johansen-Schätzer). Hätte sie länger überlegt, hätte sie die korrekte Antwort später gegeben, oder gar nicht, dann wäre die Reaktionszeit unendlich lang (was auch nach  $t$  liegt). Dieser Umstand wird von der Kaplan-Meier-Methode, wie auch von der verwendeten Form des Aalen-Johansen-Schätzers berücksichtigt, indem theoriegeleitet ein langsamerer Wert anstelle der Fehlreaktion gesetzt wird. Rein rechnerisch imputiert der RS auch einen Wert anstelle der Fehlreaktionen, nämlich den Mittelwert der richtigen Reaktionen. Auch im RS wird nicht nur mit „realen“ Reaktionszeiten gearbeitet. Warum dieser Wert zur Korrektur genommen wird, wird weder begründet, noch ist es ein valideres Vorgehen. KM und AJ bieten nicht nur theoretisch validere Alternativen, sondern ermöglichen tendenziell genauere Schätzungen der Reliabilität.

Es gibt drei Limitationen der gefundenen Ergebnisse: a) Die Stichproben sind nicht optimal, b) die Testverfahren sind nicht die besten Verfahren, c) es wurde nur die Reliabilität und nicht auch die Validität untersucht.

Die Stichproben und Testverfahren wurden aufgrund ihrer Verfügbarkeit verwendet. Die Anforderungen an die Stichproben waren eine Stichprobengröße von  $N \geq 100$ , eine Fehlerrate über 10% und dass die Datensätze Rohdaten waren. Für das Erheben einer optimalen Stichprobe fehlte die Zeit. Die Testverfahren sind demnach auch nicht aus theoretischen Überlegungen heraus gewählt worden, sondern die Stichprobeneigenschaften bestimmten die Wahl der Verfahren.

Eine optimale Stichprobe zur Untersuchung der Wirksamkeit von Fehler- und Ratekorrekturen sollte heterogen und möglichst groß sein. Zudem sollte eine Fehlerrate von über 10% per Testperson vorliegen. Andernfalls unterscheiden KM und AJ in ihrer Schätzung kaum vom RS.

Anstelle der Reliabilität hätte auch die Verbesserung der Validität bei Anwendung von Fehler- und Ratekorrekturen untersucht werden können. Für eine Untersuchung der Validität könnte die Korrelation zwischen einem externen Kriterium und den Reaktionszeiten der TAP verwendet werden.

Wenn KM und AJ zu einer genaueren Schätzung der „wahren“ Performanz führen, sollte die Korrelation unter Anwendung der KM und des AJ höher sein im Vergleich zum RS. Es gibt insgesamt wenig Validitätsuntersuchungen zur TAP. Validitätsuntersuchungen mit Einsatz von Fehler- und Ratekorrekturen könnten für weitere Untersuchungen von Interesse sein.

## Referenzen

- Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics*, 3, 15-27.
- Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics*, 6, 534-545.
- Aalen, O., & Johansen, S. (1978). An empirical matrix for non-homogenous Markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5, 141-150.
- Borgan, Ø. (2015). Aalen–Johansen Estimator. *Wiley StatsRef: Statistics Reference Online*. doi: 10.1002/9781118445112
- Efron, B. (1976). The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 4, 831–853. Verfügbar unter <https://projecteuclid.org/euclid.bsmsp/1200513831>
- Eriksen, C. W. (1966). Independence of successive inputs and uncorrelated error in visual form perception. *Journal of Experimental Psychology*, 72, 26-35.
- Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant target. *Perception & Psychophysics*, 44, 191-193. doi: 10.3758/BF03208712
- Földényi, M., Giovanoli, A., Tagwerker-Neuenschwander, F., Schallberger, U., & Steinhausen, H-C. (2006). Reliabilität und Retest-Stabilität der Testleistungen von 7-10 jährigen Kindern in der computerunterstützten TAP. *Zeitschrift für Neuropsychologie*, 11, 1–11. <https://doi.org/10.1024//1016-264X.11.1.1>.
- Gondan, M., & Heckel, A. (2008). Testing the race inequality: A simple correction procedure for fast guesses. *Journal of Mathematical Psychology*, 52, 322-325.
- Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention Perception Psychophysics*, 78, 723-735.
- Heathcote, A., Brown, S. D., & Wagenmakers, E-J. (2014). The falsifiability of actual decision-making models. *Psychological Review*, 121, 676-678. doi: 10.1037/a0037771
- Johansen, S. (1978). The Product Limit Estimator as Maximum Likelihood Estimator. *Scandinavian Journal of Statistics*, 5, 195-199.
- Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121, 1-32. doi: 10.1037/a0034190



- Kaplan, E. I., & Meier, P., (1958). Nonparametric estimation from incomplete observations. *American Statistical Association Journal*, 53, 457–481.
- Link, S. W. (1982). Correcting response measures for guessing and partial information. *Psychological Bulletin*, 92, 469-486.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, 14, 247-279.
- Ollman, R. (1966). Fast Guesses in choice reaction time. *Psychonomics Science*, 6, 155-156.
- Pachella, R.G. (1974). *The interpretation of Reaction Time in Information Processing Research*. Springfield: National Technical Information Center.
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Renner, G., Lessing, T., Krampen, G., & Irblich, D. (2012). Reliabilität und Retest-Stabilität der „Testbatterie zur Aufmerksamkeitsprüfung für Kinder“ (KITAP) bei 6- bis 7-jährigen Kindern. *Zeitschrift für Neuropsychologie*, 23, 27 – 36.
- Shentu, Y., & Xie, M. (2010). A note on dichotomization of continuous response variable in the presence of contamination and model misspecification. *Statistics in Medicine*, 29, 2200-2214. doi: 10.1002/sim.3966
- Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *TRENDS in Neurosciences*, 27, 161-168. doi: 10.1016/j.tins.2004.01.006
- Van der Heijden, A. H. C., Schreuder, R., Maris, L., & Neerinx, M. (1984). Some evidence for correlated separate activation in a simple letter-detection task. *Perception & Psychophysics*, 36, 577-585.
- Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, 7, 208-256.
- Yellott, J. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 8, 159-199. doi: 10.1016/0022-2496(71)90011-3
- Ziegler, A., Lange, S., & Bender, R., (2002). Überlebenszeitanalyse: Eigenschaften und Kaplan-Meier Methode. *Deutsche Medizinische Wochenschrift*, 127, T14–T16. doi: 10.1055/s-2002-32819

Zimmermann, P., & Fimm, B. (2002). A test battery for attentional performance. In P. Zimmermann, & M. Leclercq (Hrsg.). *Applied Neuropsychology of Attention: Theory, Diagnosis and Rehabilitation*, (110-151). London: Psychology Press.

Zimmermann, P., & Fimm, B. (2016). *TAP. Test of Attentional Performance. Version 2.3*. Herzogenrath: Psytest.

## Abbildungsverzeichnis

Abbildung 1 zeigt das theoretische Verhältnis von Geschwindigkeit zu Genauigkeit. Bei steigender Geschwindigkeit der mittleren Reaktionen sinkt die Genauigkeit auf Zufallsniveau herab. Bei steigender Genauigkeit verlangsamt sich die Geschwindigkeit der mittleren Reaktionen. Die Kurve nähert sich einer 100% Genauigkeit asymptotisch an, d.h. eine höhere Genauigkeit kann nur durch starke Reduktion der Geschwindigkeit erreicht werden. ....	7
Abbildung 2 beispielhafte Darstellung eines Wettlaufmodells. Der obere Prozess stellt eine informierte Reaktion dar. Der untere Prozess eine uninformierte Reaktion. Die Informationsakkumulation für eine informierte Reaktion erreicht zuerst die Barriere und würde eine Reaktion auslösen. ....	11
Abbildung 3 Schematische Darstellung zwei distinkter mentaler Prozesse. Ausgehend von einem Stimulus gelangt Testperson S mit einer Wahrscheinlichkeit von $g$ in Zustand A, oder mit einer Wahrscheinlichkeit von $1-g$ in Zustand B. In Zustand A sollte eine 100% richtige Reaktion erfolgen. In Zustand B kann sie mit der Wahrscheinlichkeit $b$ eine richtige Reaktion treffen, oder mit $1-b$ eine Fehlreaktion machen. ....	16
Abbildung 4 Graphische Darstellung der Wahrscheinlichkeit unter $t$ zu reagieren. AB sind redundante Durchgänge, A oder B sind jeweils Durchgänge mit einem Zielreiz. Mit * markierte Durchgänge inkludieren 10% lucky guesses. Während ohne lucky guesses ein Koaktivierungsmodell gelten würde ( $P(RT < t, AB)$ größer als $P(RT < t, A+B)$ ), müsste aufgrund der lucky guesses ein Wettlaufmodell angenommen werden. ....	18
Abbildung 5 Angenommene Abfolge der mittleren Reaktionszeiten.....	19
Abbildung 6 Aufgabenbildschirm des visual search task. In dieser 5 x 5 Matrix muss nach einem Quadrat, welches nach oben hin geöffnet ist (hier in der Mitte) gesucht werden. Die Distraktoren bestehen ebenso aus geöffneten Quadraten. ....	20
Abbildung 7 Aufgabenbildschirm der divided attention der visuellen Zielreize. Die 6-8 Kreuze können zufällig an jeder Stelle der Matrix auftauchen. Bilden sie, wie in diesem Fall, ein Quadrat, so muss per Tastendruck reagiert werden.....	22
Abbildung 8 Alternativer Aufgabenbildschirm der divided attention. Die vier Abbildungen zeigen die vier möglichen Stimuli. Zielreize sind hierbei lediglich 01 und 10. Jeder Stimuli taucht in der Mitte des Bildschirmes auf. ....	22

Abbildung 9 Stimuli des Subtest go-nogo in der ersten Bedingung. Bei einem x muss per Tastendruck reagiert werden. Bei einem + soll nicht reagiert werden. ....23

Abbildung 10 Die 5 Stimuli aus dem Subtest go-nogo. 2 von ihnen werden vorab als mögliche Zielreize definiert. Nur auf diese soll per Tastendruck reagiert werden.....24

## Tabellenverzeichnis

Tabelle 1 Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit *)	6
Tabelle 2. Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktionen mit *)	8
Tabelle 3 Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit *)	11
Tabelle 4 Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit *)	14
Tabelle 5 Fiktive Reaktionszeitdaten mit ihrer mittleren Reaktionszeit (Fehlreaktion mit *)	17
Tabelle 6. Vierfelderschema der Ratedurchgänge im visual search task	21
Tabelle 7. Retestrelabilität per RS, KM, $KM_{KTT}$ , AJ und $AJ_{KTT}$ sowie die Genauigkeit zu T1, T2	26
Tabelle 8. Retestrelabilität per RS, KM, $KM_{KTT}$ , AJ und $AJ_{KTT}$ sowie die Genauigkeit zu T1, T2	27
Tabelle 9. Split-Half-Reliabilität RS, KM, $KM_{KTT}$ , AJ und $AJ_{KTT}$ sowie die Genauigkeit zu T1, T2.	27
Tabelle 10. Split-Half-Reliabilität von RS, KM, $KM_{KTT}$ , AJ und $AJ_{KTT}$ sowie die Genauigkeiten zu T1 und T2	28

## Abkürzungsverzeichnis

AJ	Aalen-Johansen Schätzer
RS	Reduced-Sample Schätzer
KM	Kaplan-Meier Methode