

Passivrauchexposition und Lungenkrebsrisiko:
Replikation und Erweiterung der Meta-Analyse
von Stayner et al. (2007)

für die

Berufsgenossenschaft Nahrungsmittel und Gaststätten

- Abschlussbericht -

Juni 2009

Projektleitung:

Prof. Dr. Michael Bosnjak

Bericht:

Dipl.-Psych. Holger Pahl, Dipl.-Psych. Robin Würfel, Prof. Dr. Michael Bosnjak

Hinweis:

Dieser Bericht stellt eine Zusammenfassung der Arbeiten von Pahl (2009) und Würfel (2009) dar, welche dem Auftraggeber mit diesem Bericht in elektronischer Form übermittelt wurden. Gemäß den Regeln guter wissenschaftlicher Praxis enthalten die Arbeiten von Pahl (2009) und Würfel (2009) neben ausführlichen Studienbeschreibungen alle verwendete Materialien (Datensätze, Codiermanual, Codierformulare, etc.).

Die vollständige Projektdokumentation kann über den folgenden URL bezogen werden:

<http://bgn-stayner-replication.project.bosnjak.eu/>

Inhaltsverzeichnis

Inhaltsverzeichnis	2
Zusammenfassung der zentralen Befunde.....	4
1 Hintergrund und Problemstellung	5
2 Untersuchungleitende Fragestellungen.....	6
3 Methode.....	12
3.1 Beschreibung der Ausgangsdatenbasis der direkten Replikation.....	12
3.2 Neucodierung der Primärstudien	16
3.3 Analysestrategien / Vorgehen.....	18
3.4 Artefaktkorrektur nach Hunter und Schmidt (2004).....	21
4 Ergebnisse.....	24
4.1 Zu Fragestellung 1: Qualität der einbezogenen Primärstudien.....	24
4.2 Replikation des Gesamteffekts anhand der im Artikel von Stayner et al. (2007) berichteten Daten.....	28
4.2.1 Schätzung des Gesamteffekts.....	28
4.2.2 Dosage-Response Analysen	29
4.2.3 Untersuchung auf selektive Veröffentlichung: Trichtergrafikanalyse, N-Q Plots, Rosenberg und Orwin.....	30
4.2.4 Leave-one-out Analysen.....	33
4.2.5 Artefaktkorrektur und Variation der Gewichtungsbasis	34
4.3 Replikation des Gesamteffekts anhand der neu codierten Primärstudien ..	34
4.3.1 Schätzung des Gesamteffekts.....	34
4.3.2 Dosage-Response Analysen	36
4.3.3 Untersuchung auf selektive Veröffentlichung: Trichtergrafikanalyse, N-Q- Plots, Rosenberg und Orwin.....	37
4.3.4 Leave-one-out Analysen.....	39
4.3.5 Artefaktkorrektur und Variation der Gewichtungsbasis	40
4.4 Auslotung des möglichen Ergebnisspektrums des Gesamteffekts unter Einbezug der im Stellschraubenmodell vorgesehenen subjektiven Entscheidungen.....	41
4.4.1 Schätzung des Gesamteffekts auf Basis der Maximalwerte	41
4.4.2 Schätzung des Gesamteffekts auf Basis der Minimalwerte.....	42

4.4.3 Zusammenfassung beider Ansätze und ausgewählte Befunde aus Pahl (2009) und Würfel (2009).....	43
5 Zusammenfassung und Bewertung.....	47
6 Literatur	49

Zusammenfassung der zentralen Befunde

Der vorliegende Bericht fasst eine Reihe eigener meta-analytischer Befunde zum Zusammenhang zwischen Passivrauchexposition am Arbeitsplatz und dem Lungenkrebsrisiko zusammen. Ausgangspunkt war eine bereits publizierte Meta-Analyse von Stayner et al. (2007), die repliziert und erweitert wurde. Es wurde der Frage nachgegangen, ob die bei Stayner et al. (2007) berichteten Befunde als zuverlässig, robust und widerlegungssicher einzustufen sind. Diese Frage muss im Hinblick auf die im Rahmen der vorgelegten Replikation und Erweiterung aus drei Gründen verneint werden.

Erstens brachte die vorliegende Replikationsstudie zu Tage, dass die bei Stayner et al. (2007) einbezogenen Primärstudien großteils von 'mittlerer' Qualität sind. Es steht zu befürchten, dass deshalb die Qualität der Befunde im Sinne einer 'garbage-in, garbage-out'-Problematik als nicht sehr hoch einzuschätzen ist. Zudem sind qualitativ hochwertigere Studien mit einer tendenziell geringeren Effektstärkenschätzung systematisch verknüpft, d.h. 'bessere' Studien suggerieren einen 'geringeren' Zusammenhang zwischen der Passivrauchexposition und dem Risiko, an Lungenkrebs zu erkranken.

Zweitens zeigte sich, dass der bei Stayner et al. (2007) berichtete Gesamteffekt in Höhe einer rund 25 prozentigen Risikoerhöhung durch Passivrauchexposition am Arbeitsplatz unter einer vergleichbar inkonsistenten methodischen Vorgehensweise zwar replizierbar ist, unter Berücksichtigung und Modellierung der getroffenen subjektiven Entscheidungen während des meta-analytischen Prozesses allerdings kaum mehr von robusten Befunden gesprochen werden kann. Nulleffekte, d.h. die Möglichkeit, dass kein Zusammenhang zwischen Passivrauchexposition am Arbeitsplatz und dem Lungenkrebsrisiko besteht, sind im Rahmen dieser Erweiterungsstudie zugrunde gelegten Modells ebenso denkbar wie eine Risikoerhöhung um rund 70 %.

Drittens hat sich gezeigt, dass die bei Stayner et al. (2007) berichtete Dosage-Response-Analyse zum Zusammenhang zwischen der Dauer der Passivrauchexposition am Arbeitsplatz in Jahren sowie dem Lungenkrebsrisiko bereits nach Ausschluss einer einzigen Studie (Reynolds et al., 1996) verschwindet, nach Herauslassen einer anderen Studie (Kreuzer et al., 2000) deutlich ansteigt.

1 Hintergrund und Problemstellung

Der vorliegende Bericht fasst meta-analytische Befunde zum Zusammenhang zwischen Passivrauchexposition am Arbeitsplatz und dem Lungenkrebsrisiko zusammen. Ausgangspunkt war eine bereits publizierte Meta-Analyse von Stayner et al. (2007), die repliziert und erweitert wurde.

Die Notwendigkeit zur Replikation ergab sich aus einer Reihe von Unstimmigkeiten in der Befunddarstellung des Ausgangsartikels, die von Bosnjak (2007) zusammengefasst wurden. Bemängelt wurde vor allem die Dokumentationsqualität der Stayner et al. (2007) Studie sowie die aus statistischer Sicht unzulässige Vermischung von Effektstärkeschätzern (Statistisches 'Fruchtsalatproblem', siehe Brüderl, 2004). Darüber hinaus wurde bei einer berichteten Dosage-Response-Analyse zum Zusammenhang zwischen der Dauer der Passivrauchexposition und dem Risiko, an Lungenkrebs zu erkranken von Bosnjak (2007) vorhergesagt, dass durch die Entfernung einer einzigen Studie (Reynolds et al., 1996) der berichtete Zusammenhang nicht mehr ermittelbar sein sollte.

Diese Unstimmigkeiten können mittels Re-Analyse des zugrunde liegenden Original-Datensatzes nicht aufgeklärt werden, da die Autoren des Originalbeitrags die Daten entgegen den einschlägigen wissenschaftlichen Gepflogenheiten nicht freigeben. Eine grundlegende Neucodierung und Re-Analyse der berücksichtigten Primärstudien war deshalb indiziert.

Das übergeordnete Ziel der Studie bestand darin, die *Zuverlässigkeit, Robustheit und Widerlegungssicherheit* der meta-analytischen Befunde bei Stayner et al. (2007) mittels moderner meta-analytischer Verfahren und Techniken zu überprüfen.

Von Seiten des Auftraggebers begründet sich das Projekt aus der für die Berufsgenossenschaft Nahrungsmittel und Gaststätten (BGN) zentralen Frage nach dem Gefährdungspotenzial von Passivrauchexpositionen am Arbeitsplatz. Auf diese Frage scheint die von Stayner et al. (2007) veröffentlichte Meta-Analyse kaum zufriedenstellende, nachvollziehbare, nahezu umfassende und vor allem widerspruchsfreie Antworten liefern zu können.

2 Untersuchungsleitende Fragestellungen

Zur Überprüfung der Zuverlässigkeit, Robustheit und Widerlegungssicherheit der meta-analytischen Befunde bei Stayner et al. (2007) wurde die Ausgangsstudie aus fünf verschiedenen Perspektiven repliziert. Die daraus abgeleiteten untersuchungsleitenden Fragestellungen werden nachfolgend skizziert (siehe auch Pahl, 2009, und Würfel, 2009).

Fragestellung 1: Qualität der einbezogenen Primärstudien?

Zunächst soll der Frage nachgegangen werden, von welcher Qualität die bei Stayner et al. (2007) synthetisierten Primärstudien sind. Als Qualitätsindikatoren dienen die einschlägigen STROBE-Kriterien (Vandenbrouke et al., 2007) sowie das aus der Psychologie stammende, von Wittmann und Matt (1986) vorgeschlagene Bewertungsschema. Neben einer reinen Beschreibung der Primärstudienbasis über Qualitätsindikatoren sollen diese mit Effektstärken (hier: Risiko an Lungenkrebs zu erkranken, formalisiert als Odds Ratio Metrik) in Beziehung gebracht werden. So kann ermittelt werden, ob und inwieweit die Studienqualität mit der Stärke des Effekts von Passivrauchen am Arbeitsplatz und dem Lungenkrebsrisiko in Beziehung steht.

Fragestellung 2: Replizierbarkeit und Robustheit des Gesamteffekts sowie der Dosage-Response-Analyse anhand der im Artikel von Stayner et al. (2007) berichteten Daten?

Im Rahmen der zweiten zu behandelnden Fragestellung sollen die im Artikel von Stayner et al. (2007, Tabelle 1, S. 2) berichteten Effektstärkenschätzungen der Primärstudien meta-analytisch synthetisiert werden. Es wird somit überprüft, ob sich der berichtete Gesamtbefund sowie die Dosage-Response-Analyse (hier: Zusammenhang zwischen der zeitlichen Exposition durch Passivrauch und Lungenkrebsrisiko) anhand der im Artikel wiedergegebenen Datenbasis replizieren lässt. Hierbei werden die von Bosnjak (2007) geäußerten Bedenken im Hinblick auf das sogenannte 'statistische Fruchtsalatproblem' (Brüderl, 2002) nicht berücksichtigt. Die Robustheit der Befunde soll mittels Sensitivitätsanalysen und Tests auf eine selektive Studienausswahl bzw. -veröffentlichung untersucht werden.

Fragestellung 3: Replizierbarkeit und Robustheit des Gesamteffekts sowie der Dosage-Response-Analyse anhand neu codierten Primärstudien?

Das analytische Vorgehen entspricht hier dem der vorhergehenden Fragestellung 2, die Datenbasis ist jedoch nicht dem Originalartikel entnommen, sondern basiert auf einer umfangreichen Neucodierung der bei Stayner et al. (2007) berücksichtigten Primärstudien. Die Regeln zur Identifikation von Effektstärken orientieren sich am im Originalartikel skizzierten Vorgehen. Mittels Fragestellung 3 können eventuelle Übertragungsfehler bei der Codierung eingegrenzt werden.

Fragestellung 4: Robustheit des Gesamteffekts durch Einbezug subjektiver Entscheidungen im meta-analytischen Prozess?

Im Rahmen der vierten Fragestellung soll untersucht werden, ob und inwieweit subjektive Entscheidungen, die während eines jeden meta-analytischen Prozesses in unterschiedlichen Phasen zu treffen sind (Bosnjak & Viechtbauer, im Druck), das Ergebnisspektrum im Falle der Stayner et al. (2007) Studie bedeutsam beeinflusst hätten. Die hier verfolgte Frage lautet:

‘Welche Zusammenhänge zwischen Passivrauchexposition am Arbeitsplatz und dem Lungenkrebsrisiko würden sich ergeben, wenn man an unterschiedlichen Entscheidungspunkten im meta-analytischen Prozess eine andere, aber dennoch inhaltlich wie auch methodisch plausible Entscheidung getroffen hätte?’

Die Frage zum Einfluss subjektiver Entscheidungen steht im Zusammenhang mit der übergeordneten Zielsetzung der Studie, die *Zuverlässigkeit, Robustheit und Widerlegungssicherheit* der meta-analytischen Befunde von Stayner et al. (2007) zu überprüfen. Dass subjektive Entscheidungen im Forschungsprozess meta-analytische Befunde beeinflussen können, ist bereits einschlägig dokumentiert (siehe hierzu Bosnjak & Viechtbauer, im Druck). Im vorliegenden Fall steht im Vordergrund, wie stark diese subjektiven Entscheidungen die Befunde im vorliegenden Fall beeinflusst haben könnten, d.h. es wird eine *Quantifizierung der potenziellen Ergebnisvariation durch subjektive Entscheidungen im Forschungsprozess angestrebt*.

Ausgangspunkt zur Bearbeitung der Fragestellung ist das im Rahmen dieses Forschungsprojekts entwickelte ‘Stellschraubenmodell’. Ziel des Modells ist die Rekonstruktion subjektiver, aber dennoch jeweils begründbarer Entscheidungen im Rahmen eines meta-analytischen Prozesses auf der Ebene der Datengenerierung und Datenanalyse (siehe Bosnjak & Viechtbauer, im Druck; Pahl, 2009, S. 33 ff.; Würfel, 2009, S. 22 ff; für eine ausführliche Beschreibung des Modells). Angewandt auf eine konkrete meta-analytische Studie erlaubt das Modell den Einfluss der Summe aller subjektiven Entscheidungen auf die Bandbreite der Befunde abzuschätzen.

Bezogen auf die Stayner et al. (2007) Meta-Analyse wird der Einfluss der folgenden ‘Stellschrauben’ auf die mögliche Bandbreite der Befunde ermittelt:

- Adjustierte versus nicht adjustierte Effektstärken: Sollen für die metaanalytische Berechnung nur Studien verwendet werden, die statistisch unkorrigierte Ergebnisse liefern oder nutzt man nur die Ergebnisse der Studien für die Analyse, die statistisch angepasst wurden oder nutzt man beide Formen zusammen innerhalb einer Metaanalyse?
- Geschlecht der Untersuchungsgruppe: Man kann Effektstärken für beide Geschlechter kombiniert berechnen, oder getrennt nach dem jeweiligen Geschlecht.
- Effektstärkendefinition: Diese Stellschraube stellt einen Sonderfall dar, da sie – im Gegensatz zu den beiden erst genannten – nicht direkt aus den einzelnen Primärstudien abgelesen werden kann. Die verwendeten Primärstudien unterscheiden sich stark darin, was jeweils unter

„Environmental Tobacco Smoke“ (ETS) Aussetzung am Arbeitsplatz verstanden wird. Deutlich machen kann man dies an den Fragen, die den jeweiligen Patienten und Kontrollpersonen gestellt wurden. Die Frage der engen Definition lautet dabei: „Sind Sie am Arbeitsplatz, und nur dort, regelmäßig Passivrauch ausgesetzt?“ Eine positive Antwort auf diese Frage geben also lediglich Personen, die wirklich nur an ihrem Arbeitsplatz regelmäßig Passivrauch ausgesetzt sind und an keinem weiteren Ort darüber hinaus (sogenannte 'EA-Gruppe', siehe Pahl, 2009, S. 65). Die Frage hinter der mittleren Definition lautet folgendermaßen: „Sind Sie am Arbeitsplatz regelmäßig Passivrauch ausgesetzt?“ Im Unterschied zu der obigen Frage unterscheidet sich die Gruppe, welche auf diese Frage mit ja antwortet dadurch, dass sie zwar auf jeden Fall regelmäßig auf der Arbeit Passivrauch ausgesetzt sind, es aber möglich ist, dass sie noch an weiteren Orten regelmäßig Passivrauch inhalieren. Ob es tatsächlich weitere Orte gibt, an denen zumindest ein gewisser Teil dieser Personen wiederholt Passivrauch ausgesetzt sind, ist anhand der Antwort auf diese Frage nicht eindeutig zu klären (EA + EOA Gruppe, siehe Pahl, 2009, S. 65). Hinter der weiten Definition steht die folgende Frage: „Sind Sie regelmäßig Passivrauch ausgesetzt?“ Diese Frage ist so offen formuliert, dass sie sowohl von Personen, die auf der Arbeit Passivrauch ausgesetzt sind, als auch von Menschen, in deren Gegenwart ausschließlich zu Hause oder an anderen Orten geraucht wird, bejaht werden kann! Es ist also durchaus möglich, dass Personen aus dieser Gruppe Passivrauch auf der Arbeit ausgesetzt waren. Ob dies aber tatsächlich der Fall ist und wenn ja, wie viele es genau sind, ist in keiner Weise erkennbar (EA + EOA + EO Gruppe, siehe Pahl, 2009, S. 65).

- Verwendung einer Subgruppe: Unter einer Subgruppe werden hier nur solche Personengruppen erfasst, die Passivrauch in einem bestimmten zeitlichen Intervall bzw. einer bestimmten Intensität ausgesetzt sind. Man kann also jeweils die einzelnen Effektstärken der Subgruppen in eine Meta-Analyse aufnehmen, oder man nimmt die Effektstärke der gesamten Studie zusammen.

Ausführliche Begründungen für die Wahl dieser Stellschrauben finden sich bei Pahl (2009, S. 63 ff) sowie Würfel (2009, S. 51 ff.).

Fragestellung 5: Robustheit des Gesamteffekts nach Artefaktkorrektur?

Im Rahmen der fünften und letzten Fragestellung sollen potenzielle Artefakte wie bei Hunter und Schmidt (2004) beschrieben berücksichtigt werden. Da dieser Ansatz außerhalb der Medizin und Epidemiologie kaum bekannt ist, wird er nachfolgend skizziert.

Hunter und Schmidt (2004) diskutieren eine Reihe von Artefakten, die Einfluss auf das Ergebnis einer Metaanalyse haben können. Bisher vor allem im Rahmen von arbeits- und organisationspsychologischen Fragestellungen angewandt soll nun versucht werden, diese Artefakte auch in der vorliegenden Arbeit zu bestimmen. Das so genannte 'Attenuationsmodell' beschreibt bei Hunter und Schmidt das zustande kommen der beobachteten Effektstärken. Dabei kommt es zu einer entweder systematischen oder unsystematischen Minderung des

wahren Effektes, wodurch es zu einer Über- oder Unterschätzung des wahren Effektes und dessen Konfidenzintervalls kommt.

Hunter und Schmidt schlagen über das sogenannte 'Disattenuationsmodell' eine Korrektur der beobachteten Effektstärken vor. Hierzu werden diese nachträglich um die für die Fragestellung bekannten Artefakte korrigiert. In einem zweiten Schritt wird auch die Gewichtung über die Teilnehmerzahl um einen sogenannten 'Gesamtinderungsfaktor' ergänzt.

Nachfolgend werden die Artefakte von Hunter und Schmidt beschrieben und ihre Anwendbarkeit auf die vorliegende Fragestellung und Primärstudiengruppe geprüft:

1. Der Stichprobenfehler ('sampling error') beschreibt einen unsystematischen Effekt der Stichprobengröße. Dieses Artefakt kann nicht für einzelne Studien korrigiert werden. Allerdings stellt die Stichprobengröße ein Haupteinflussfaktor für die Größe dieses Artefakts dar, also wird ihm durch die Gewichtung, basierend auf der Stichprobengröße, Rechnung getragen.
2. Der Messfehler in der abhängigen und der unabhängigen Variable beschreibt eine unsystematische Minderung des Effekts. Hunter und Schmidt schlagen eine Korrektur für jede Effektstärke vor. Hierzu werden allerdings die Reliabilitäten für die Messinstrumente in den verwendeten Studien benötigt. Da diese in unserer Studiengruppe nicht vorliegen und auch nicht genügend Angaben gegeben werden, um die Reliabilität nachträglich beurteilen zu können, wird stattdessen um einen konstanten Wert korrigiert. Die Reliabilität des ärzteurteiltes wird von Ashton (2000) beschrieben, die Reliabilität von Patientenangaben zu ETS Aussetzung am Arbeitsplatz von Pron et al. (1988). Die Ergebnisse dieser Studien sind streitbar, andere Studien kommen zum Teil zu erheblich anderen Ergebnissen (bspw. Huang, 2008). Sie bilden aber eine gute Grundlage um den Effekt der Korrektur zu zeigen.
3. Bei der künstlichen Dichotomisierung von Daten handelt es sich um eine Umwandlung einer kontinuierlichen Variable in eine bipolare, bspw. durch einen Mediansplit. Dieser ebenfalls unsystematische Effekt ist bei Fragestellungen der A&O-Psychologie häufig, wenn z.B. Bewerber für eine Stelle aufgrund eines Intelligenztestes als geeignet (1) oder ungeeignet (0) eingeteilt werden. Gehen in die folgende Analyse nur diese Werte ein, wird die kontinuierliche Information, die ein Intelligenztest bietet, nicht genutzt. Da eine Einschränkung der Varianz einer Variable auch immer eine Minderung der mit ihr zusammenhängenden Korrelation zur Folge hat, wird auch hier der Effekt unterschätzt. Ob sich dieses Artefakt auf die vorliegende Fragestellung anwenden lässt ist allerdings fraglich, da auf abhängiger Variablenseite, also dem Vorhandensein einer Krebserkrankung, keine künstliche Dichotomisierung besteht. Auf den gegebenen unabhängigen Variablen, z.B. der Dauer der Krebsaussetzung, sind in den vorliegenden Primärstudien nicht genug Informationen angegeben, um diese Korrektur durchzuführen. Auch ist die Art der angewendeten Analysen zu unterschiedlich, so dass keine allgemeine Korrektur wie bei der Reliabilitätskorrektur vorgenommen werden kann.

4. Beim Artefakt der Variabilitätsbereichsverzerrung (range variation) geht es um eine Unter- bzw. Überschätzung der Varianz bei der unabhängigen Variable. Wie bei der künstlichen Dichotomisierung hat diese Varianz Einfluss auf die daraus folgende Korrelation. Bei den vorliegenden Daten wurde geprüft, ob es eine Einschränkung von Patienten gab, die befragt wurden. So weichen z.B. die Ausschlusskriterien in den Primärstudien, die einen Nichtraucher definieren, voneinander ab. Auch hier sind diese Informationen nicht häufig genug gegeben, um eine Korrektur vorzunehmen. Bei der abhängigen Variable wird ein ähnlicher Effekt 'Attrition' (Abnahme) genannt. Dieser Effekt drückt eher den Schwund durch Prozesse während der Untersuchung aus. Bei der vorliegenden Fragestellung könnte es sich hierbei um verstorbene oder aus dem Krankenhaus entlassene Patienten handeln. Eine Angabe, die nur sehr inkonsistent in der vorliegenden Literatur berichtet wird und daher noch nicht korrigierbar ist.
5. Fehler in der Konstruktvalidität bei der abhängigen und unabhängigen Variable drücken im Gegensatz zu der Reliabilität systematische Fehler bei der Operationalisierung von Fragestellungen aus. Die Ungenauigkeit der Messung wird ähnlich korrigiert wie bei der Reliabilität, um den Faktor der Abweichung des gemessenen Wertes von dem wahren Konstrukt. Hierzu sind genaue Kenntnisse der Probleme bei der Messung der vorliegenden Konstrukte nötig. Da bei unserem Projekt weder Mediziner noch Epidemiologen mit dem nötigen Fachwissen beteiligt waren, wurde dieses Artefakt bei der vorliegenden Arbeit nicht berücksichtigt. Indirekt kann aber die Beurteilung der Studienqualität als Indikator für eine gelungene Konstruktvalidität aufgefasst werden, welche als Moderator in diesem Artikel mit untersucht wurde (siehe 4.1)
6. Fehler in den Daten und bei der Übertragung von Daten, wie sie wahrscheinlich in jeder wissenschaftlichen Arbeit vorkommen, sind besonders schwer zu berücksichtigen. Schon in dem Artikel von Stayner et al. (2007) wurden von den Autoren dieser Arbeit einige potenzielle Übertragungs- und Berechnungsfehler entdeckt, allerdings lässt sich keine Aussage über die Menge dieser Fehler in den Primärstudien machen. Da es leider auch in der Literatur keine Schätzungen gibt, wie hoch die Fehlerrate durchschnittlich ist, wird dieser Artefakt hier nicht berücksichtigt.
7. Bei dem letzten Artefakt wird der Einfluss abweichenden Vorgehens der Primärstudien bei der Untersuchung der Fragestellung untersucht. Hierbei handelt es sich um Drittvariablen, die Einfluss auf die Effektstärke haben aber bei der Untersuchung nicht mit berücksichtigt wurden. Viele dieser Moderatoren wurden von den Autoren der vorliegenden Arbeit angedacht (siehe Codiersheet und Codiermanual), auch hier ist aber die Datengrundlage nicht ausreichend, um Korrekturen vorzunehmen. Eine weitere Möglichkeit besteht in der nachträglichen Korrektur von möglichen Moderatoren, allerdings wurde aus Zeitgründen auf diese Möglichkeit verzichtet.

Zu erwähnen ist noch eine Daumenregel für die Homogenitätsrechnung. Die 75%-Regel besagt, dass wenn mindestens 75% der Varianz in den beobachteten, korrigierten Effektstärken auf die (korrigierte) Stichprobenfehlervarianz zurückzuführen ist, von Homogenität ausgegangen werden kann. Die restliche

Variabilität ($\leq 25\%$) geht dann auf unerklärte Artefakte und unsystematische Unterschiede zwischen den Studien ($>$ Modell zufallsvariabler Effekte!) zurück. Hunter und Schmidt (2004) erwähnen zwar auch einen Q-Test zur Homogenitätstestung, schätzen dessen Aussagekraft aufgrund geringer Power jedoch sehr kritisch ein.

3 Methode

Die Arbeit von Stayner et al. (2007) wird in verschiedenen Schritten repliziert. Zunächst wird die Replikation der Meta-Analyse anhand der im Artikel von Stayner und Kollegen berichteten Primärstudienresultate vorgenommen. Im Anschluss werden die darin verwendeten Primärstudien neu codiert und mit den daraus gewonnenen Daten weitere Meta-Analysen durchgeführt.

3.1 Beschreibung der Ausgangsdatenbasis der direkten Replikation

Die Ausgangsdatenbasis für die direkte Replikation bezieht sich ausschließlich auf die in der Arbeit von Stayner et al. (2007) berichteten Ergebnisse der dort berücksichtigten Primärstudien. Darüber hinaus werden in diesem Schritt keine Hinweise aus den verwendeten Primärstudien verwendet. Die Autoren beschreiben darin, dass ihre Analyse zu den Auswirkungen von Passivrauch am Arbeitsplatz auf das Lungenkrebsrisiko als eine Erweiterung bestehender Metaanalysen, einmal um neuere Primärbefunde und weiterhin um neue Arten von Berechnungen mit den Daten, zu verstehen ist. Grundlage für die Primärstudienbeschaffung sind ältere Metaanalysen zu dem gleichen Thema sowie eine Literaturrecherche via MedLine und Embase am 01. Januar 2003.

Insgesamt werden 22 Studien in die Metaanalyse aufgenommen, aus denen im ersten Reanalyseschritt 25 Effektstärken verwendet werden (Tabelle 1). Wann immer es möglich ist werden adjustierte Werte verwendet, in einigen Fällen muss jedoch auf nicht adjustierte Werte zurückgegriffen werden. Neben den relativen Risiken sind die Konfidenzintervalle (CI) und die wichtigsten Studiencharakteristiken (*key study variables*) codiert und anschließend in verschiedenen Analyseformen ausgewertet worden (siehe auch Pahl, 2009, S. 27, für weitere Details).

Tabelle 1 Deskriptive Daten der 22 Primärstudien sowie die dazugehörigen Effektstärken und Konfidenzintervalle (Stayner et al., 2007, S. 546)

Reference ^a	Location	Time Period	Gender	No. Cases	Covariate Adjustments ^b	Exposure Period	Histologic Confirmation	RR (95% CI)
Kabat et al. ¹⁸ (w)	USA	1971-1980	Women	53	None	Current	No	0.7 (0.3, 1.5) ^c
Kabat et al. ¹⁸ (m)	USA	1971-1980	Men	25	None	Lifetime	No	3.3 (1.0, 10.6) ^c
Koo et al. ¹⁹	Hong Kong	1981-1983	Women	88	None	Lifetime	Yes	1.2 (0.5, 3.0) ^f
Garfinkel et al. ²⁰	USA	1971-1981	Women	76	a	Lifetime	Yes	0.9 (0.7, 1.2)
Wu et al. ²¹	Los Angeles	1981-1982	Women	29	a	Lifetime	No	1.3 (0.5, 3.3) ^d
Lee et al. ²² (w)	England	1979-1982	Women	15	None	NR	No	0.6 (0.2, 2.3) ^f
Lee et al. ²² (m)	England	1979-1982	Men	10	None	NR	No	1.6 (0.4, 6.6) ^f
Shimizu et al. ²³	Japan	1982-1985	Women	90	a	Current	Yes	1.2 (0.6, 2.6) ^{e,g}
Kalandidi et al. ²⁴	Greece	1987-1989	Women	89	None	Lifetime	No	1.4 (0.8, 2.5) ^f
Wu-Williams ²⁵	China	1985-1987	Women	415	a	Lifetime	No	1.2 (0.9, 1.6) ^h
Kabat et al. ²⁶ (w)	USA	1983-1990	Women	58	a	Lifetime	Yes	1.2 (0.6, 2.1)
Kabat et al. ²⁶ (m)	USA	1983-1990	Men	41	a	Lifetime	Yes	1.0 (0.5, 2.1)
Reynolds et al. ²⁷	USA	1986-1990	Women	528	a,d,s,r,o	Lifetime	Yes	1.6 (1.2, 2.0)
Schwartz et al. ²⁸	USA	1984-1987	Both	257	a,r	NR	No	1.5 (1.0, 2.2)
Sun et al. ²⁹	China	NR	Women	230	a	NR	Yes	1.4 (0.9, 2.0)
Wang et al. ³⁰	China	1992-1994	Women	135	None	NR	No	0.9 (0.5, 1.8)
Boffetta et al. ³¹	Europe	1988-1994	Both	650	a	Lifetime	No	1.2 (0.9, 1.5)
Boffetta et al. ³²	Europe	1994-1996	Both	70	a	Lifetime	No	1.5 (0.8, 3.0) ^d
Zaridze et al. ³³	Russia	NR	Women	189	a	Current	Yes	0.9 (0.6, 1.4)
Rapiti et al. ³⁴	India	1991-1992	Both	58	a	Lifetime	Yes	1.1 (0.3, 4.1)
Zhong et al. ¹⁵	China	1992-1994	Women	504	a	Lifetime	No	1.7 (1.3, 2.3)
Kreuzer et al. ³⁵	Germany	1990-1996	Both	123 ^k	a,r	Lifetime	Yes	1.1 (0.7, 1.7)
Lee et al. ³⁶	Taiwan	1992-1998	Women	268	a,o	Lifetime	Yes	1.5 (0.5, 2.4)
Wang et al. ³⁷	China	1994-1998	Both	233	a	Lifetime	No	1.6 (0.7, 3.3)
Johnson et al. ³⁸	Canada	1994-1997	Women	71	None	Lifetime	Yes	1.3 (0.4, 4.0)

Note. NR = not reported.

^aStudies also used to plot Figures 1-3.

^bCovariates that were adjusted for in the analysis: a = age; r = race; d = diet; s = exposure to environmental tobacco smoke from spouse; o = occupational exposure to other carcinogens.

^cWe estimated the 95% CI on the basis of the results presented in the article in which the results appear.

^dResults are for adenocarcinoma of lung only.

^eThe 95% CI was not reported. It was estimated with the average standard error taken from Kalandidi et al.²⁴ and Nyberg et al.³³ because all 3 studies had similar numbers of lung cancer cases.

^fThe reported result was 1.1 (95% CI = 0.9, 1.6); Wells et al. reported the correct estimates.¹⁵

^gSome of the cases and controls in Kreuzer et al.³⁵ were part of another study included in this table (Boffetta et al.³¹). The results given here are based on those cases and controls that were not part of the Boffetta study (M. K., written communication, 2002).

Für die Neuberechnung der Gesamteffektstärke müssen die angegebenen Risiken zunächst logarithmisiert werden, um eine Annäherung an die Normalverteilung zu erreichen und somit eine symmetrische Größe zu erhalten. Anschließend muss die Residualvarianz bestimmt werden, was mittels der angegebenen Konfidenzintervalle, bzw. deren logarithmisierte Form, möglich ist. Mit den somit vorliegenden Werten und dem Kehrwert der Varianz als Gewichtungsfaktor können die Gesamteffektstärken berechnet werden. Die einzelnen Formeln zur Berechnung des mittleren Effekts werden beispielsweise in der Arbeit von Schulze (2004) dargestellt (siehe auch Pahl, 2009, S. 47).

Daneben kommen bei Stayner et al. (2007) zwei Dosage-Response-Analysen zum Einsatz, und zwar (a) durch die Berücksichtigung lediglich der Personen mit der höchsten Intensität der ETS-Aussetzung am Arbeitsplatz und (b) die zeitliche Dauer der ETS-Aussetzung am Arbeitsplatz.

(zu a) Die Berechnung des Risikos an Lungenkrebs zu erkranken, wenn Personen in extremer Form ETS am Arbeitsplatz ausgesetzt sind, berechnen die Autoren anhand von 7 Studien. Die Werte der größten Intensität dieser Studien werden zu einem Effekt aggregiert. In der dazugehörigen Tabelle 2 aus der Arbeit von Stayner und Kollegen (2007, S. 547) werden alle zur Berechnung herangezogenen Werte mit weiteren Informationen abgebildet (siehe auch Pahl, 2009, S. 28):

Tabelle 2 Effektstärken und Konfidenzintervalle der Gruppen mit den höchsten ETS-Intensitäten (Stayner et al., 2007, S. 547)

Reference	Gender Included in Study	Exposure Measure ^a	RR (95% CI)
Boffetta et al. ³¹	Both	≥ 89 level × hours per day × years ^b	2.07 (1.33, 3.21)
Johnson et al. ³⁸	Women	≥ 64 smokers × years	1.58 (0.6, 4.0)
Kabat et al. ¹⁸	Men	Smokers × hours per week × years ^c	1.21 (0.47, 3.13)
Kabat et al. ¹⁸	Women	Smokers × hours per week × years ^c	1.35 (0.64, 2.84)
Kalandidi et al. ²⁴	Women	Duration × number coworkers ^d	1.08 (0.24, 4.87)
Kreuzer et al. ³⁵	Both	> 100.6 level × hours per day × years ^b	2.64 (1.07, 6.54) ^e
Lee et al. ³⁶	Men	Average to a lot	0.46 (0.05, 4.65) ^f
Zhong et al. ¹⁵	Women	≥ 4 coworkers smoked	3.0 (1.8, 4.9)
Meta-analysis			
Fixed effects			2.01 (1.55, 2.60)
Mixed effects			2.01 (1.33, 2.60)

^aThe measure of exposure used to categorize workers varied from study to study. For studies that presented more than 1 measure, preference was given to exposure measures that reflected both intensity and duration (i.e., cumulative exposures).

^bThe total number of years of exposure weighted for the number of hours of exposure per day and for a subjective index of level of smokiness at the workplace (1 = very smoky, 0.5 = fairly smoky, and 0.2 = a little smoky).

^cThe highest tertile of exposure was compared with the lowest tertile. The actual values of the tertiles were not presented in the article.

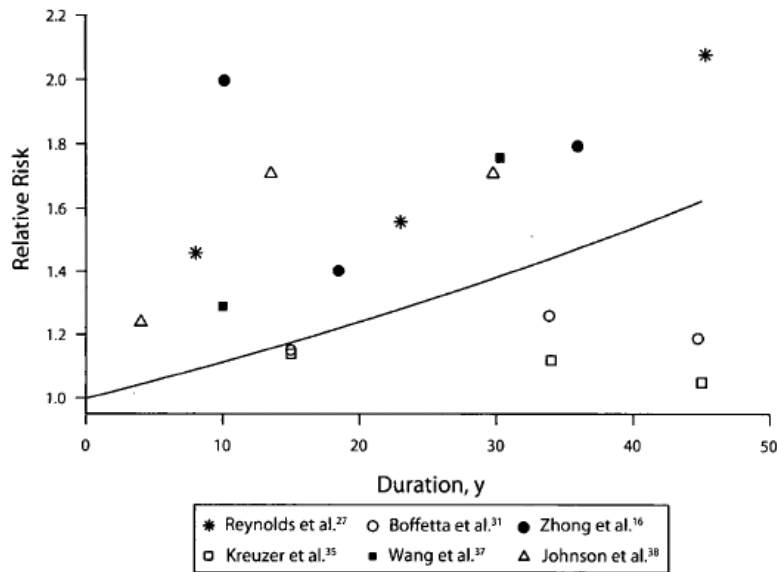
^dThe results are for a comparison between the highest and lowest quartiles of "the time-weighted sum of exposure at work, the exposure being based on the number of smokers among people working in the same closed space." The units of these quartiles are not presented in the article.

^eResults are from an analysis that excluded cases and controls that were in the analysis by Boffetta et al.,³¹ which was not presented in the original analysis.

^fCrude results not adjusted for any risk factors.

Die eigentliche Berechnung der Gesamteffektstärke für die Subgruppe der höchste Intensitäten erfolgt analog wie oben beschrieben, allerdings nur mit den in Tabelle 2 abgebildeten Werten.

(zu b) Zusätzlich berechnen Stayner und Kollegen mit 6 Studien unter der Bezeichnung „Meta-Regression“ eine Analyse zur Vorhersage des Risikos durch die Länge der Passivrauchaussetzung, indem sie eine lineare Regression über die 6 Studien analysieren. Unter dem Begriff Meta-Regression werden im eigentlichen Sinne Subgruppenanalysen verstanden, um die Heterogenität von Effekten zwischen Studien aufzudecken (Morton, Adams, Suttorp & Shekelle, 2004). Darüber hinaus soll anhand der gegebenen Informationen das Risiko an Lungenkrebs zu erkranken geschätzt werden. Als Beispiel wurde eine Aussetzung von 45 Jahren angenommen. Um für die Abhängigkeit zwischen mehreren Datenpunkten einer Studie zu korrigieren, wird eine Methode von Greenland und Longnecker (1992) angewendet. Alle verwendbaren Studien und deren Effektstärken werden in Abbildung 1 dargestellt:



Note. The diagonal line is the fitted fixed model effect of duration.

Abbildung 1: Regression zur Vorhersage des Lungenkrebsrisikos über die Dauer der Aussetzung von ETS (Stayner et al., 2007, S. 548)

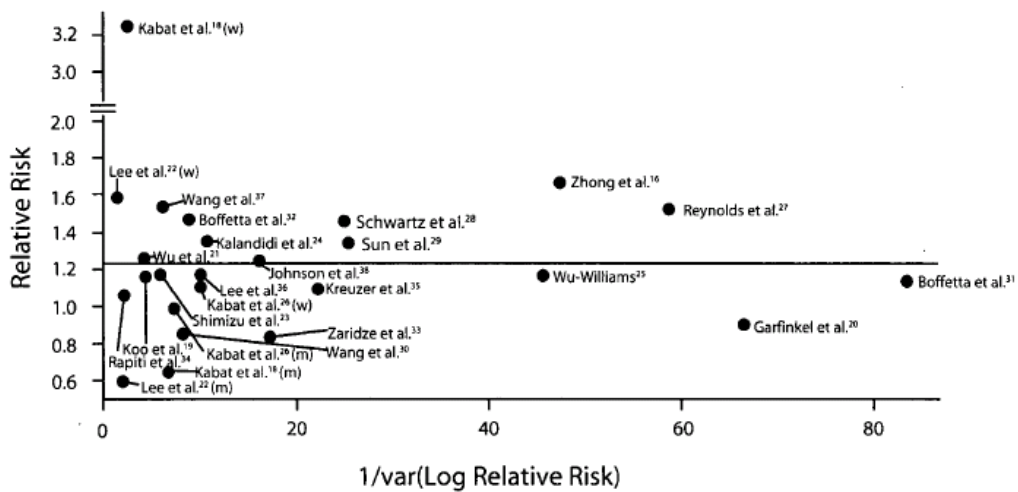
Dieser Replikationsschritt ist alleine anhand der Informationen aus der Arbeit von Stayner und Kollegen nicht durchführbar. Stattdessen mussten die Datenpunkte aus Abbildung 1 vermessen werden, um eine annähernde Replikation zu bewerkstelligen. Dies wurde zum einen per Hand und zum anderen mit dem Programm g3data¹ gemacht. Da darüber hinaus keine weiteren quantitativen Angaben aus der Arbeit von Stayner und Kollegen über die Regression zu entnehmen sind, werden alle Datenpunkte ungewichtet aufgenommen und mittels SPSS eine Meta-Regression wie bei Stayner et al. (2007) berechnet. Es wird zunächst für jede der dabei verwendeten Primärstudien, über die jeweils gegebenen Punkte aus der oben abgebildeten Figur von Stayner et al. (2007), ein eigener b_k -Wert samt zugehörigem Standardfehler berechnet. Die einzelnen Werte werden anschließend gepoolt zu einem einzelnen Wert zusammengefasst. Die gepoolte Standardabweichung wird auf einen Wert von .005 festgesetzt. Mit den erhaltenen Werten kann das Risiko an Lungenkrebs zu erkranken berechnet werden, abhängig von der Aussetzungsdauer (siehe auch Pahl, 2009, S. 52).

Um den Einfluss jeder einzelnen Studie zu bewerten, berechnen Stayner et al. (2007) mehrere Sensitivitätsanalysen meist in Form von „leave-one-out“ Analysen, aber auch über die Verwendung ausschließlich adjustierter Studien. Die Gegenanalyse, welche nur unkorrigierte Werte berücksichtigt, wurde nicht berichtet. Die Berechnung der einzelnen Gesamteffektstärken erfolgt analog wie oben beschrieben.

Zur Untersuchung, ob ein Publication Bias bei der Auswahl der verwendeten Primärstudien vorliegt, wählen Stayner et al. (2007) die grafische Darstellung

¹ <http://www.frantz.fi/software/g3data.php>

eines Funnel Plots. In diesem bilden sie die Effektstärken über die jeweilige Größe im Verhältnis zum jeweiligen Kehrwert der Varianz ab, was in der vorliegenden Replikation ebenfalls durchgeführt wird. Der Funnel Plot ist in der nachfolgenden Abbildung 2abgetragen (siehe auch Pahl, 2009, S. 29).



Note. The horizontal line is the meta-analysis relative risk estimate, which was 1.24.

Abbildung 2: Funnel Plot zur Abschätzung einer möglichen Verzerrung anhand einer selektiven Veröffentlichungspraxis (Stayner et al., 2007, S. 549)

3.2 Neucodierung der Primärstudien

Für den zweiten Replikationsteil, der Replikation anhand einer Neucodierung aller in Stayner et al. (2007) berichteten Primärstudien, mussten zunächst die Inhalte und entsprechenden Angaben aller verwendeten Primärstudien codiert werden. In der vorliegenden Studie wurde ein eigens auf die Fragestellung zugeschnittenes Codiermanual entwickelt und angewendet, welches sich in Appendix G der Diplomarbeit von Holger Pahl (2009) finden lässt. Dieses war die Basis für die Erfassung der einzelnen Primärstudien. Die Gestaltung des Codiermanuals erfolgte in mehreren Schritten. In vorangegangenen Überlegungen wurde festgehalten, dass für die Codierung verschiedene Themenblöcke von Bedeutung sind. Diese bestehen aus Variablen zur allgemeinen Beschreibung der jeweiligen Studie, Variablen rund um die Berechnungen der Effektstärken und Variablen zur Studienqualität. Letztendlich resultierten somit für die vorliegende Arbeit vier Codierblöcke: (1) Designvariablen, (2) Variablen zur Qualität nach der Cochrane Collaboration, (3) Variablen rund um die Effektstärkenberechnung und (4) Variablen zur Qualität nach psychologischen Beurteilungskriterien (angelehnt an Wittmann & Matt, 1986). Die unten stehende Abbildung 3 zeigt den Aufbau der Codierung schematisch:



Abbildung 3: Grundlagenmodell hinter der Codierung (Würfel, 2009, S. 36)

Das Codiermanual beschreibt dabei jede einzelne Variable näher und enthält weiterhin Angaben darüber, welche Ausprägungen die jeweiligen Variablen einnehmen können. In das mit Excel erstellte Codierformular wurden hingegen die einzelnen Eintragungen vorgenommen.

Der *erste Block* beinhaltet Variablen, die die einzelnen Studien und ihre Inhalte beschreiben. Beispiele sind grundlegende Dinge wie die Namen der Autoren oder das Publikationsjahr, aber auch inhaltliche Angaben wie die Anzahl der befragten Personen und die Krankheiten, welche in die Untersuchung mit einbezogen wurden. Die meisten Variablen in Block 1 sind nominalskaliert.

Der *zweite Block* des Codierschemas richtet sich nach der Arbeit von Vandembroucke et al. (2007) zur Verbesserung der Berichterstattung bei Beobachtungsstudien in der Epidemiologie, was mit STROBE abgekürzt wird. In diesem Artikel wird eine Checkliste beschrieben, mit deren Hilfe Autoren zukünftiger epidemiologischer Arbeiten die Berichtsqualität erhöhen sollen. Anhand dieser Checkliste werden die verwendeten Primärstudien der Metaanalyse von Stayner et al. (2007) überprüft. Studien, welche besonders viele der geforderten Punkte erfüllen, werden als qualitativ hochwertig interpretiert. Dieser Block codiert die einzelnen Variablen somit jeweils mit 1 für *ja* und 0 für *nein*. Im Anschluss kann so ein Summenscore über die gesamten dichotomen Items gebildet werden, der eine Aussage über die Qualität der einzelnen Studie nach den STROBE Kriterien liefert. Items dieses Blocks sind ausschließlich nominal skaliert.

Der *dritte Block* enthält alle zur Effektstärkenberechnung notwendigen Variablen und Angaben der einzelnen Primärstudien. Beispielsvariablen sind die Verteilung der Personenanzahlen in den Vier-Felder-Tafeln, sofern diese angegeben sind oder die berichteten Effektstärkenwerte in den Primärstudien. Auch die selbst berechneten, unkorrigierten Effektstärken, sind in diesem Block enthalten. Weitere Angaben, wie z.B. das dazugehörige Konfidenzintervall, finden sich ebenfalls im dritten Block. Die meisten Variablen in diesem Block sind metrisch.

Wie bereits der zweite Block, so enthält auch Block vier Variablen zur Bewertung der Qualität jeder einzelnen Studie. In diesem Fall beziehen sich die Variablen aber auf psychologische Beurteilungskriterien, beispielsweise auf die externe Validität oder die Konstruktvalidität. Grobe Vorlage für die Erstellung dieses Codierblocks war eine Metaanalyse von Wittmann und Matt (1986) über die Effektivität von Psychotherapie, allerdings konnten viele Variablen nicht übernommen werden, da sich die Studiengrundlage stark unterscheidet. Die Variablen im vierten Block liegen hauptsächlich der Ordinalskala zugrunde. Um auch über diese Variablen ein Summenscore der Qualität berechnen zu können, werden die Variablen aus Block 4 jedoch nachträglich dichotom umcodiert. Diese umcodierten Variablen gehen auch in die Berechnung der Intercoder-Reliabilität ein (siehe auch Pahl, 2009, S. 43).

Zur Bestimmung der Intercoder-Reliabilität hat sich der Ansatz von Krippendorff (2004) durchgesetzt (Bosnjak & Viechtbauer, in Druck). Dieses Maß, bekannt als Krippendorffs Alpha, setzt die beobachtete Abweichung zwischen den Codierern ins Verhältnis zu einer zufällig erwarteten Abweichung. Zur Berechnung der Intercoder-Reliabilität wurden zunächst drei der Studien zufällig ausgewählt, die von allen drei Codierern bearbeitet wurden. Allgemein bewegt sich Krippendorffs Alpha zwischen Werten von 0 und 1, wobei größere Werte für eine höhere Übereinstimmung stehen. Beim Abgleich der Codierungen mittels dem Makro-Programm von Hayes (Hayes & Krippendorff, 2007) in Verbindung mit SPSS ergab sich dabei ein Alpha-Werte von .84 über alle drei Codierer. Die Berechnung erfolgte lediglich über die Codierblöcke 1, 2 und 4, das Skalenniveau wurde auf nominal festgelegt. Der dritte Block wurde von allen Codierern gemeinsam bearbeitet und somit nicht in die Intercoder-Reliabilitätsberechnung aufgenommen. Anschließend wurden die restlichen 19 Primärstudien unter den drei Bearbeitern aufgeteilt und codiert (siehe Pahl, 2009, S. 46).

3.3 Analysestrategien / Vorgehen

Um die oben beschriebenen Analysen mit den Daten der Neucodierung replizieren zu können, werden zunächst die einzelnen Effektstärken der Primärstudien ohne Adjustierung ermittelt, indem die angegebenen Vierfeldertafeln verwendet werden (Pahl, 2009, S. 49). Daneben ist es in einigen Fällen möglich, Effektstärken getrennt nach Geschlechtern zu berechnen, selbst wenn nicht immer die vollständige Vierfeldertafel eines Geschlechts in der Primärstudie abgebildet ist. Dies ist der Fall, wenn in der Primärstudie Angaben zu Effektstärken und den dazugehörigen Vierfeldertafeln für beide Geschlechter zusammen und für ein Geschlecht alleine angegeben sind. Weiterhin gibt es verschiedene Definitionen hinter der Frage, ob die beteiligten Personen auf der

Arbeit Passivrauch ausgesetzt sind oder nicht. Durch die Angaben zu Verteilungen von Subgruppen – beispielsweise Personen, die Passivrauch in einem bestimmten zeitlichen Intervall an ausgesetzt sind – und die oben genannten Punkte, ist es bei dem Großteil der verwendeten Primärstudien möglich, deutlich mehr als nur eine passende Effektstärke mit dazugehörigem Konfidenzintervall zu berechnen.

Nach der Beschreibung des Vorgehens in der Metaanalyse von Stayner und Kollegen gilt es im Anschluss für die Berechnung der Gesamteffektstärke über alle Studien, Effektstärken in die Berechnungen aufzunehmen, die „adjusted for confounders“ (Stayner et al., 2007, S. 545) angegeben werden. Dies ist bei dem Großteil der Studien der Fall, in einigen Fällen gibt es jedoch keine Anpassung der Rohdaten in den Primärstudien. Daneben gilt, dass bevorzugt Effektstärken aufgenommen werden, die Ergebnisse getrennt nach Geschlecht berichten. Andere Arten von Subgruppen, wie oben beschrieben, werden für die Berechnung der Gesamteffektstärken nicht berücksichtigt. Das Vorgehen bei der Auswahl der jeweiligen Effektstärken erfolgt in diesem Schritt deshalb nach folgendem Muster:

- Es wird adjustierten Werten der Vorzug vor nicht-adjustierten Werten gegeben.
- Die Verwendung von Effektstärken, die der mittleren Definition von Passivrauch am Arbeitsplatz folgen wird bevorzugt verwendet; wo dies nicht möglich ist, wird die enge Definition verwendet.
- Es werden bevorzugt Werte für beide Geschlechter getrennt aufgenommen, ansonsten jedoch keine weiteren Subgruppenergebnisse verwendet.

Die Methodiken hinter der Berechnung der Gesamteffektstärke und der Homogenitätstests sind methodisch identisch mit der direkten Replikation. Trotz der unterschiedlichen Aussagen des Fixed-Effect Modells und des Random-Effect Modells, werden auch bei der Replikation anhand der Daten aus den Primärstudien beide Modelle berechnet, um eine möglichst vollständige Replikation abbilden zu können. Innerhalb der Moderatoranalyse wird die Liste der key study variables erneut untersucht. Allerdings sieht diese Liste im Falle der Neucodierung etwas anders aus als im Original von Stayner et al. (2007). So werden die Autoren der Primärstudien an dieser Stelle in Gruppen zusammengefasst, wenn sie an mehreren der ausgewählten Primärstudien beteiligt waren bzw. mit den gleichen Autoren anderer Studien zusammengearbeitet haben. Nach dieser Prozedur ergeben sich dabei drei Gruppen (Gruppe Boffetta, Gruppe Wu und Gruppe Kabat). Weiterhin werden innerhalb der Primärstudien einige Kovariaten aufgeführt über welche die Ergebnisse statistisch korrigiert wurden, welche im Original vom Stayner und Kollegen jedoch nicht genannt werden. Dies sind beispielsweise Anpassungen an die Ausbildung der Versuchspersonen oder deren Konsum von Gemüse. Neben diesen Unterschieden wird für die Moderatoranalyse anhand der Neucodierung die Variable Exposure Period nicht erhoben, da diese in den Primärstudien nicht genau genug herausgearbeitet ist und nicht nachvollzogen werden konnte, wie Stayner et al. (2007) auf die Daten in ihrer Studie gekommen sind. Die

Moderatorberechnungen selbst folgen der gleichen Methodik wie unter der direkten Replikation beschrieben (siehe Pahl, 2009, S. 56-58).

Für die Dosage-Response-Analyse zur Intensität werden alle Effektstärken der einzelnen Primärstudien zusammengetragen, die Angaben zur Subgruppe mit der höchsten Aussetzungsintensität an Passivrauch am Arbeitsplatz treffen. Entsprechend dem Vorgehen von Stayner et al. (2007) werden, sofern in einer Primärstudie mehrere Ergebnisse zu Messungen von ETS Aussetzungen vorhanden sind, Ergebnisse zu „cumulative exposure rather than intensity of exposure“ (Stayner et al., 2007, S. 546) herangezogen. Die Auswahl der Effektstärke erfolgt nach dem gleichen Muster wie bei der Berechnung der Gesamteffektstärke anhand der Neucodierung beschrieben steht. Weiterhin wird auch in diesem Teil eine Homogenitätsanalyse durch den Q-Test durchgeführt.

Auch die oben beschriebene Meta-Regression wird anhand der neucodierten Daten repliziert, indem das gleiche methodische Vorgehen wie bei Stayner et al. (2007) gewählt wird. In den meisten Fällen werden innerhalb der Primärstudien Zeitintervalle angegeben, denen Personen Passivrauch am Arbeitsplatz ausgesetzt sind. Stayner und Kollegen geben in ihrer Arbeit an, dass sie für die Regressionsberechnungen jeweils die Mittelpunkte dieser Intervalle verwenden. Die Ausnahme bilden Zeitangaben mit offenem Ende, hier wird der angegebene Zeitwert mit 1.5 multipliziert und verwendet. Dieser Wert wird auf höchstens 45 Jahre gesetzt, selbst wenn der errechnete Wert größer ausfallen sollte. Dieses Vorgehen wird in dieser Replikation erneut gewählt. Daneben wird zusätzlich ein Homogenitätstest anhand des Q-Tests über die verwendeten Studien berechnet. Für die Durchführung der Metaregression wird auf einen Artikel von Greenland und Longnecker (1992) verwiesen, um die Korrelation der Datenpunkte innerhalb der verwendeten Studien zu berücksichtigen. Dieses Vorgehen wird mithilfe der Informationen aus den vorliegenden Primärstudien wiederholt. Die Auswahl der zu verwendeten Effektstärken versucht dabei wieder dem Beispiel der Original-Metaanalyse zu folgen und sieht in diesem Fall folgendermaßen aus:

- Es ist adjustierten Werten der Vorzug vor nicht-adjustierten Werten zu geben.
- Es werden ausschließlich Effektstärken in die Regression mit aufgenommen, die der mittleren Definition von Passivrauch am Arbeitsplatz folgen.
- Es werden jeweils kombinierte Werte für beide Geschlechter zusammen aufgenommen.

Die Studie von Kreuzer, Krauss, Kreienbrock, Jockel & Wichmann (2000) stellt dabei eine Besonderheit dar: Die Dauer von ETS am Arbeitsplatz wird in der Studie von Kreuzer und Kollegen, im Gegensatz zu den anderen Primärstudien, in Stunden und nicht in Jahren angegeben. Um die Stundenangaben mit den Jahresangaben der anderen Studien vergleichen zu können, sind die Angaben dieser Studie, die in Deutschland entstanden ist, nach folgendem Prinzip von Stunden in Jahre umgerechnet worden: Ausgehend von der Annahme, dass ein Arbeitstag etwa 8 Stunden hat, eine Arbeitswoche in der Regel 5 Tage beträgt, und ein Kalenderjahr, nach Abzug des Urlaubs (≈ 30 Tage = 6 Arbeitswochen)

etwa 46 Arbeitswochen beinhaltet, berechnet sich ein Wert von $46 * 8 * 5 = 1840$ Arbeitstunden pro Jahr. Wenn Kreuzer et al. (2000) beispielsweise von 29000 Stunden sprechen, entspricht dies nach oben beschriebener Formel etwa einer Angabe von $29000 / 1840 = 15.76$ Jahren.

Eine zusätzliche Besonderheit der Studie von Kreuzer et al. (2000) ist, dass bei der Berechnung der korrigierten Effektstärken in der Studie anders verfahren worden ist, als in den anderen Primärstudien. Während in der Regel eine bestimmte Untergruppe mit einer gewissen Aussetzungsdauer von Passivrauch auf der Arbeit mit einer Untergruppe, die keinen Passivrauch auf der Arbeitsstelle ausgesetzt ist, verglichen wird, so wird bei Kreuzer et al. (2000) zum Vergleich eine Untergruppe gewählt, die ebenfalls Personen enthält, die zumindest in geringem Maße ETS am Arbeitsplatz ausgesetzt waren. Da in der Studie von Kreuzer et al. (2000) jedoch auch Angaben von den Verteilungen der Personen enthalten sind, an deren Arbeitsplatz es keinen Passivrauch gibt, wird in den eigenen Berechnungen diese Gruppe als Vergleichsgruppe gewählt, um das jeweilige OR zu berechnen.

Entsprechend der oben beschriebenen Vorgabe von Stayner et al. (2007), werden anschließend die b-Gewichte über die Regressionspunkte der einzelnen Primärstudien mittels der Methode von Greenland und Longnecker (1992) korrigiert. Die dort beschriebene Methode berücksichtigt die Korrelation zwischen Schätzungen innerhalb jeder Studie (Greenland & Longnecker, 1992). Für die Berechnung wird ein Makro-Programm für SAS verwendet. Anschließend wird ein gepooltes b-Gewicht einschließlich des zugehörigen gepoolten Standardfehlers berechnet und so eine lineare Regression bestimmt. Über diese kann ein OR für 45 Jahre ETS am Arbeitsplatz in Bezug auf das Lungenkrebsrisiko einschließlich Konfidenzintervalls bestimmt werden. Auch ein Homogenitätstest wird über die Verwendung des Q-Tests erneut durchgeführt, indem die korrigierten b-Gewichte zusammen mit dem korrigierten Kehrwert der Varianz in einem Fixed-Effects Modell gefittet werden (siehe Pahl, 2009, S. 59-61).

3.4 Artefaktkorrektur nach Hunter und Schmidt (2004)

Im folgenden Abschnitt wird das Vorgehen bei der Artefaktkorrektur beschrieben. Es wird nur die Korrektur der Reliabilität auf unabhängiger und abhängiger Variablenseite eingegangen, da die anderen Korrekturen bei der vorliegenden Studiengruppe nicht anwendbar sind.

Da in den Primärartikeln keine Angaben zu den Artefakten der Reliabilitätskorrektur auf anhängiger und unabhängiger Variablenseite gegeben sind, wird für die Reliabilität von ärztlichen Urteilen und für die Zuverlässigkeit von Patientenangaben zum Thema ETS am Arbeitsplatz je ein konstanter Wert übernommen.

In der Arbeit von Hunter und Schmidt werden im Wesentlichen Korrelationen berücksichtigt. Die für die für die Artefaktkorrekturen benötigten Effektstärken müssen deswegen von Odds Ratios in Korrelationen und nach der Korrektur wieder zurück berechnet werden.

Das genaue Vorgehen der Artefaktkorrektur beinhaltet einige Schritte:

1. Die Umrechnung der OR Werte in unkorrigierte Korrelationen

$$r_{0i} = \cos \left(\frac{180^\circ}{1 + \sqrt{\frac{ad}{bc}}} \right) \quad (1)$$

2. Die Berechnung des mit N gewichteten $\bar{r}_0 = \frac{\sum N_i * r_{0i}}{\sum N_i}$ (2)

3. Die unkorrigierte Stichprobenfehlervarianz $Var(e_{0i}) = \frac{(1 - \bar{r}_0^2)^2}{N_i - 1}$ (3)

4. Die Bestimmung des Disattenuationsfaktors a aus den Reliabilitäten $a = \sqrt{r_{xx} * r_{yy}}$ (4)

5. Die Bestimmung des Attenuationsfaktors $A = \frac{r_o}{r_c}$ (5)

6. Die Berechnung der korrigierten Korrelationen pro Studie $r_{Ci} = \frac{r_{0i}}{a}$ (6) sowie

$$\text{der korrigierten Stichprobenfehlervarianz Pro Studie } Var(e_{Ci}) = \frac{Var(e_{0i})}{A^2} \quad (7)$$

7. Bei der Gewichtung w_i wird zusätzlich zur Stichprobengröße auch der Attenuationsfaktor berücksichtigt $w_i = N_i A_i^2$ (8)

Aus diesen Grunddaten lassen sich nun die gewichteten und korrigierten Metaparameter errechnen:

8. Der korrigierte Gesamteffekt (random Model) $\bar{r}_C = \frac{\sum w_i * r_{Ci}}{\sum w_i}$ (9) mit einem

$$\text{Konfidenzintervall von } CI = \bar{r}_C \pm 1.96 \sqrt{\frac{Var(r_C)}{k}} \quad (10) \text{ bei k Studien}$$

9. Die gewichtete, korrigierte Varianz der Korrelationen

$$Var(r_C) = \frac{\sum w_i [r_{Ci} - \bar{r}_C]^2}{\sum w_i} \quad (11)$$

10. Im letzten Schritt müssen die Parameter für die Homogenitätsschätzung berechnet werden, der durchschnittliche, korrigierte Stichprobenfehler

$$Ave(Var(e_{Ci})) = \frac{\sum w_i * Var(e_{Ci})}{\sum w_i} \quad (12), \text{ welche für die Homogenitätsschätzung}$$

durch $Var(r_C)$ geteilt wird

11. Die Metaparameter können nun in Odd Ratios zurückgerechnet werden

$$OR_c = \left(\frac{180^\circ}{\cos^{-1}(\bar{r}_c)} - 1 \right)^2 \quad (13)$$

Die Stayner et al. (2007) zu Grunde liegende Studiengruppe beinhaltet noch ein weiteres Problem. Fünf der Studien, die in die Metaanalyse eingehen, geben keine Größe der Stichprobe an (Wu (1985), Shimizu (1988), Sun (1996), Rapiti (1999), Wang (2000)). Stayner et al. (2007) löste dieses Problem, indem die Gewichtung direkt über die Varianzen der Studien berechnet wurde (siehe 3.1). Dieses Vorgehen ist durch die Kombination der Stichprobengröße N mit dem quadrierten kombinierten Attenuationsfaktor (siehe Formeln (8) und (9)) hier nicht möglich. Im Folgenden werden die Studien deshalb ausgeschlossen und die korrigierten Ergebnisse werden mit einer unkorrigierten Metaanalyse verglichen, bei der die Studien ebenfalls ausgeschlossen werden.

Die Ergebnisse der Artefaktkorrektur sind im Abschnitt 4.3.5 zu finden. Da Stayner und Kollegen keine Stichprobengröße in ihrer Arbeit angeben ist auf Basis dieser Datengrundlage keine Korrektur nach Hunter und Schmidt für die erste Replikationsphase zu berechnen (siehe 4.2.5).

4 Ergebnisse

4.1 Zu Fragestellung 1: Qualität der einbezogenen Primärstudien

Nach den STROBE Items (Vandenbroucke et al., 2007) gibt es dabei eine Spanne von 5 (für die Studie von Reynolds et al., 1996) bis zu 35 Punkten (für die Studie von Boffetta et al., 1998). Die theoretisch erreichbare Höchstpunktzahl beträgt in der vorliegenden Arbeit 40 Punkte. Diese orientieren sich an den 22 Punkten und Unterpunkten des Originalartikels, zusätzlich wurden in der vorliegenden Arbeit einige Items extra für die Kranken- und Gesundengruppe getrennt behandelt. In Abbildung 4 sind die Verteilungen der Qualitätsbewertungen aller Primärstudien abgetragen:

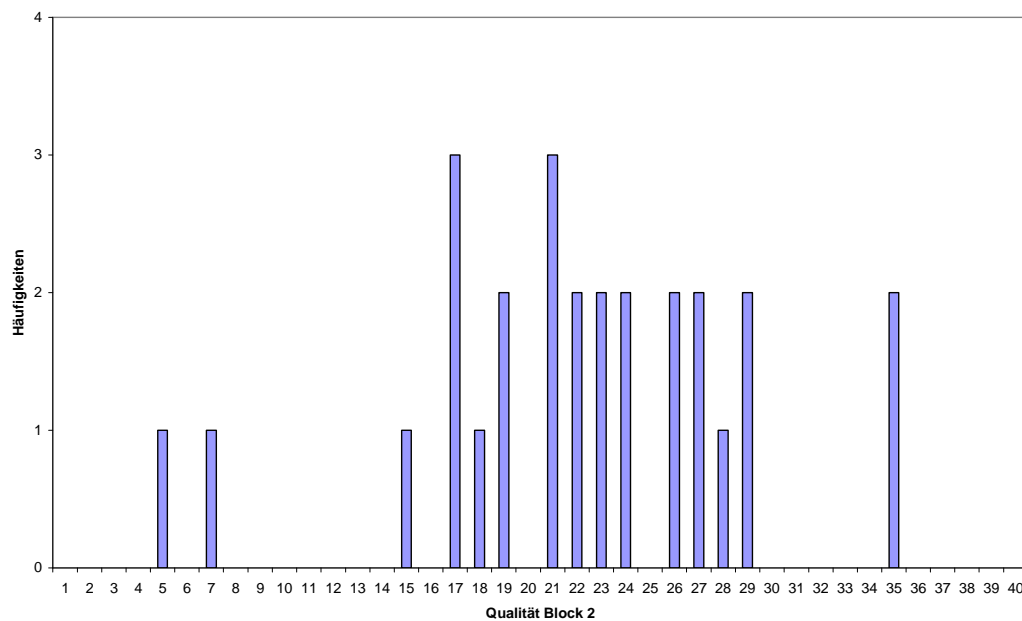


Abbildung 4: Histogramm über die Häufigkeitsverteilung der Qualität nach Block 2 über die verwendeten Primärstudien und deren Effektstärken

Anhand der Berechnung einer Moderatoranalyse durch eine Regression mit der Qualität als UV und der Effektstärke als AV soll überprüft werden, ob die Qualität einen signifikanten Einfluss auf die Effektstärke ausübt (siehe Abbildung 5).

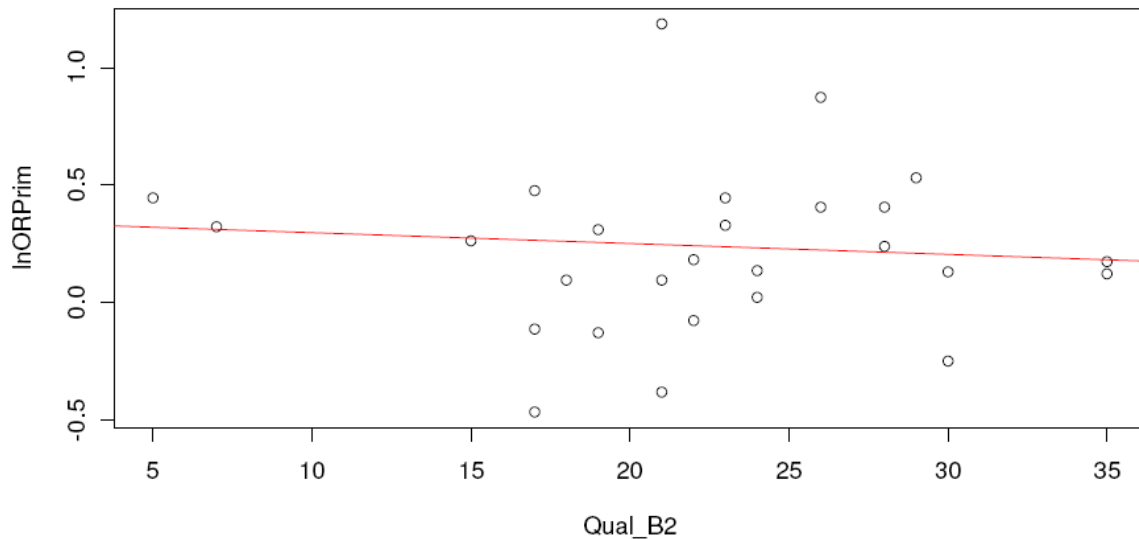


Abbildung 5: Verteilung der Qualität über alle Effektstärken der Primärstudien nach den STROBE-Items von Vandembroucke et al. (2007)

Die Punktzahlen der einzelnen Primärstudien, angewendet als Moderator über die Daten zur Berechnung der Gesamteffektstärke von Stayner et al. (2007), erreicht keine Signifikanz, weder unter Verwendung des Fixed-Effects Modells mit Moderatoren ($Q = 1.34$, $p = .25$), noch über das Mixed-Effects Modell ($Q = 0.78$, $p = .38$). Auch über die neucodierten Daten zur Berechnung der Gesamteffektstärke ergibt sich keine Signifikanz durch die Qualität nach STROBE, sowohl nach dem Fixed-Effects Modell mit Moderatoren ($Q = 1.02$, $p = .31$), als auch nach dem Mixed-Effects Modell ($Q = 0.80$, $p = .37$).

Schließt man die beiden Outlier aus Abbildung 5 aus der Regressionsberechnung über die neucodierten Daten aus, ergibt dies trotzdem erneut keine statistische Signifikanz, weder mit dem Fixed-Effects Modell mit Moderatoren noch mit dem Mixed-Effects Modell ($Q = 1.13$, $p = .29$). Der Verlauf ist in Abbildung 6 zu betrachten.

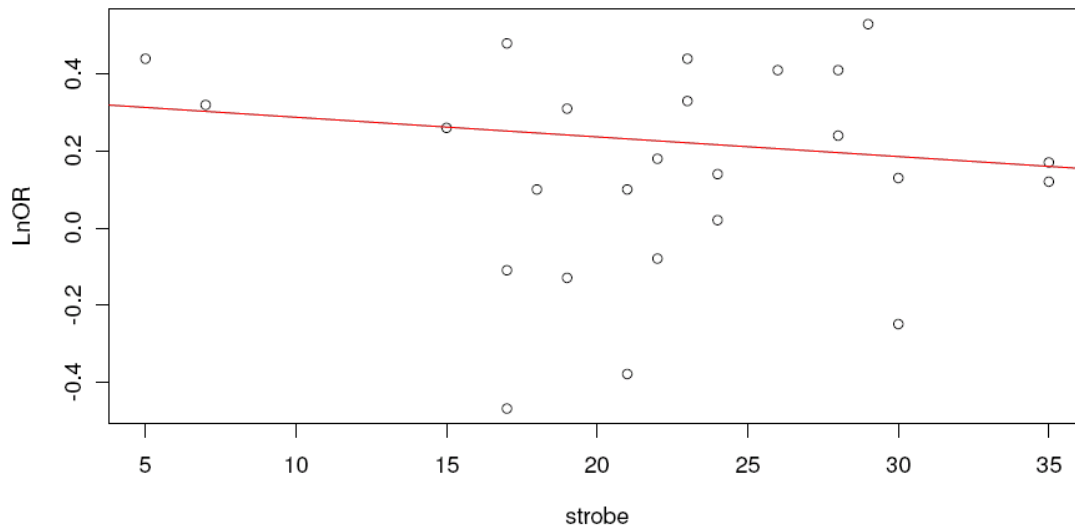


Abbildung 6: Verteilung der Qualität über alle Effektstärken der Primärstudien nach den STROBE-Items von Vandembroucke et al.(2007), ohne die Outlier aus Abbildung 5

Der zweite Teil der Moderatoranalyse durch Qualitätsindikatoren erfolgt über die Variablen aus Block 4 der Codierung. Demnach kann eine Studie maximal 10 Punkte erhalten (Würfel 2009, S. 57). In Abbildung 7 sind die Verteilungen der Qualität nach Block 4 über alle Studien dargestellt:

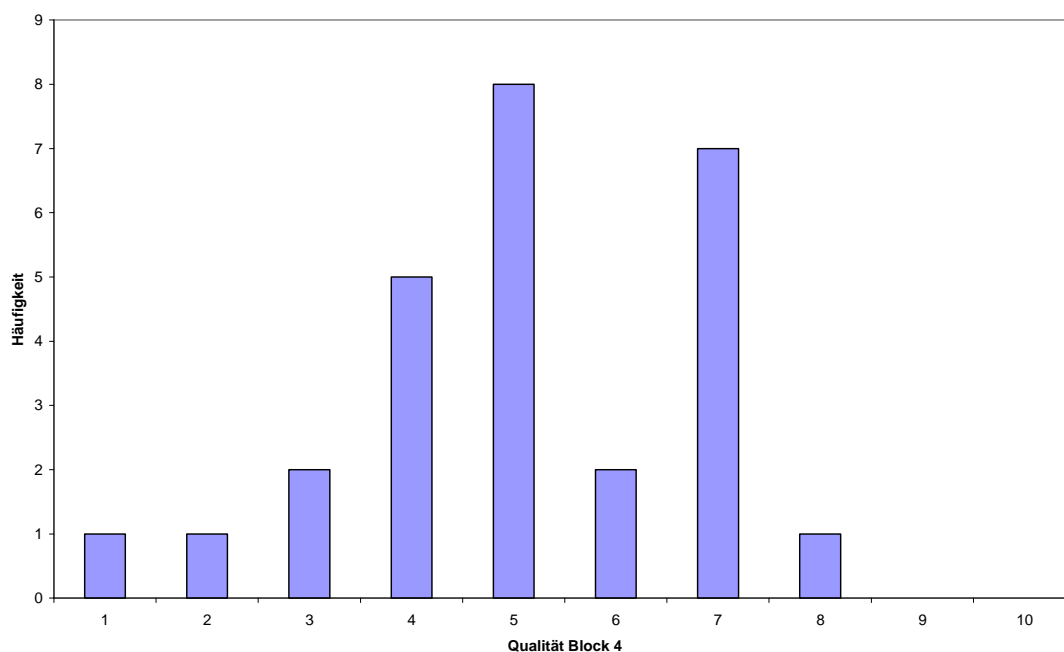


Abbildung 7: Histogramm über die Häufigkeitsverteilung der Qualität nach Block 4 über die verwendeten Primärstudien und deren Effektstärken

Den höchsten Wert der Primärstudien erhält die Studie von Lee et al. (2000) mit 8 Punkten, die Studie mit dem niedrigsten Wert stammt von Reynolds et al. (1996) mit einem Punkt. Wird eine Moderatoranalyse durch eine gewichtete Regression über die Items aus dem vierten Block der Codierung berechnet, so

ergibt sich angewendet auf die Effektstärken zur Berechnung der Gesamteffektstärke von Stayner et al. (2007) sowohl nach Fixed-Effects Modell mit Moderatoren, als auch nach dem Mixed-Effects Modell, ein signifikantes Ergebnis ($Q = 4.34$, $p = .04$). Die entsprechenden Daten der Neucodierung ergeben hingegen kein signifikantes Ergebnis, weder nach Fixed-Effects noch nach Mixed-Effects Modell ($Q = 3.09$, $p = .08$). Der beschriebene Zusammenhang ist in Abbildung 8 visualisiert, weitere Details finden sich in Pahl (2009, S. 90).

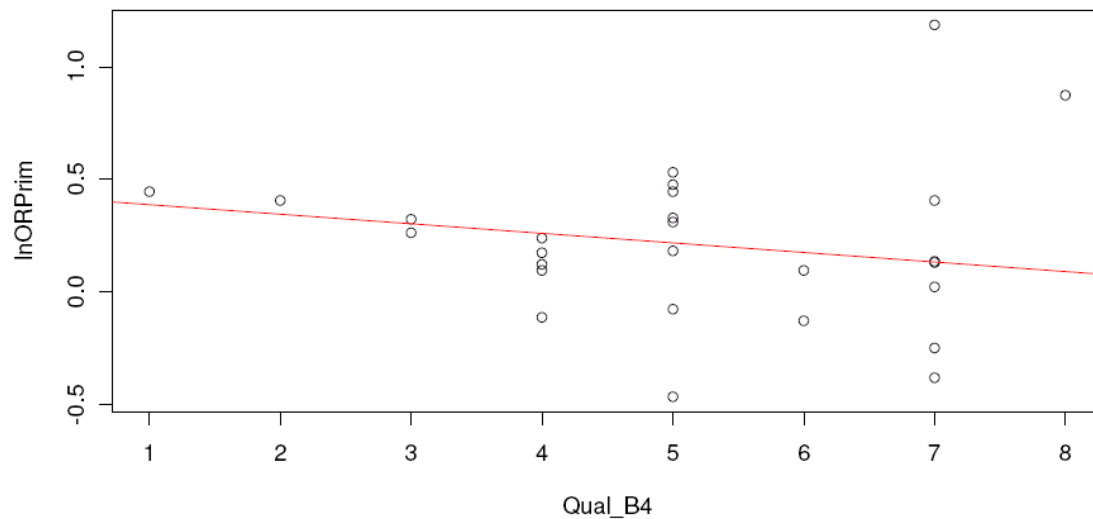


Abbildung 8: Verteilung der Qualität über alle Effektstärken der Primärstudien nach psychologischen Kriterien aus Block 4

Werden dabei jedoch die beiden Outlier aus Abbildung 8 aus der Regressionsberechnung ausgeschlossen, so führt dies zu einem signifikanten Ergebnis, sowohl nach dem Fixed-Effects Modell mit Moderatoren als auch nach dem Mixed-Effects Modell ($Q = 6.38$, $p = .01$). Der entsprechende Verlauf ist in Abbildung 9 dargestellt.

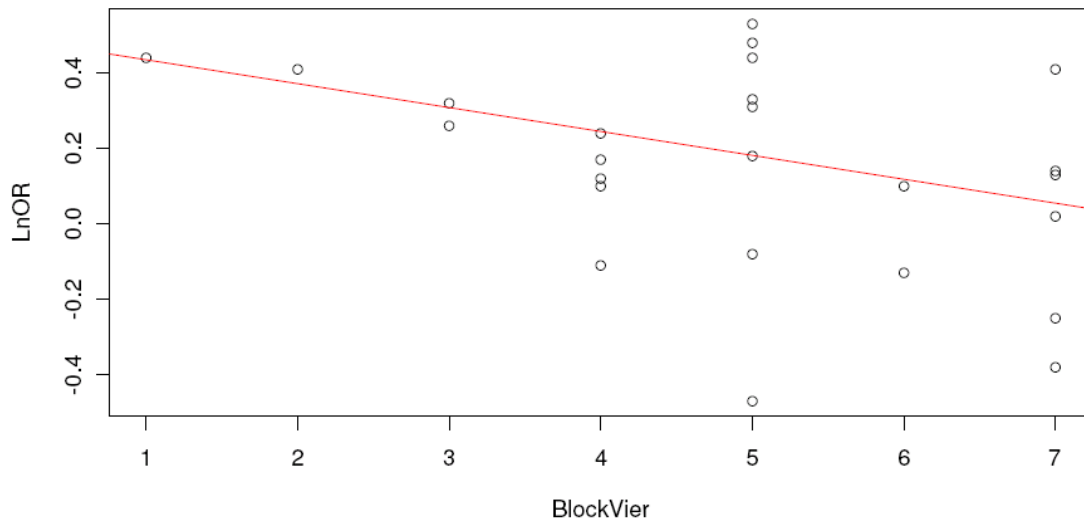


Abbildung 9: Verteilung der Qualität über alle Effektstärken der Primärstudien nach psychologischen Kriterien aus Block 4, ohne die Outlier aus Abbildung 8

4.2 Replikation des Gesamteffekts anhand der im Artikel von Stayner et al. (2007) berichteten Daten

4.2.1 Schätzung des Gesamteffekts

Nach eigenen Berechnungen über die 22 verwendeten Primärstudien und deren 25 Effektstärken, die auch von Stayner et al. (2007) verwendet wurden, erhält man nach dem Fixed-Effects Modell eine Gesamteffektstärke von $OR = 1.28$ (CI: 1.17; 1.40) und nach dem Random-Effects Modell eine Größe von $OR = 1.27$ (CI: 1.16; 1.40). Die geringen Abweichungen zwischen den hier berichteten Effektstärken und denen von Stayner und Kollegen lassen sich möglicher Weise auf Rundungsungenauigkeit zurückführen, da diese die jeweiligen Effektstärken der einzelnen Studien lediglich auf zwei Nachkommastellen angegeben haben. Die Unterschiede bei den Konfidenzintervallen resultieren daneben aus der Verwendung unterschiedlicher Programme (SAS vs. SPSS). Der Q-Test zeigt, dass die verwendeten Effektgrößen als homogen betrachtet werden können ($Q = 26.23$, $df = 24$, $p = .34$), die Varianzkomponente der zufälligen Effekte τ^2 beträgt demnach .0050. Die durchgeführten Moderatoranalysen über die einzelnen key study variables ergeben sowohl nach Fixed-Effects Modell als auch nach dem Random-Effects Modell insgesamt 6 Moderatoren, die bei einem α -Niveau von 5% signifikante Ergebnisse auf dem Q-Test ergeben. Diese heterogenen Moderatoren sind: der Zeitraum ($Q = 7.11$, $p = .01$), die Anzahl der Fälle ($Q = 4.44$, $p = .04$), Anpassung Diät ($Q = 4.59$, $p = .03$), Anpassung Partner ($Q = 4.59$, $p = .03$), Anpassung an andere krebserregende Stoffe auf der Arbeit ($Q = 5.27$, $p = .02$) und der Expositionszeitraum. Während es bei den anderen aufgeführten Moderatoren keinen Unterschied zwischen Fixed-Effects mit Moderatoren Modell und dem Mixed-Effects Modell gibt, ist dies beim Moderator zum Expositionszeitraum anders. Nach dem Fixed-Effects mit Moderatoren Modell ergibt der Q-Test einen Wert von 4.07 ($p = .0436$) und nach dem Mixed-Effects Modell einen Wert von 3.88 ($p = .0488$), trotzdem fällt der Moderator in beiden Fällen signifikant aus. Die einzelnen Effektstärken, sowie die ermittelte Gesamteffektstärke nach dem Fixed-Effects und dem Random-Effects Modell

werden in einem Forest Plot in der nachfolgenden Abbildung abgetragen (siehe Pahl, 2009, S. 74 und Appendix A-a).

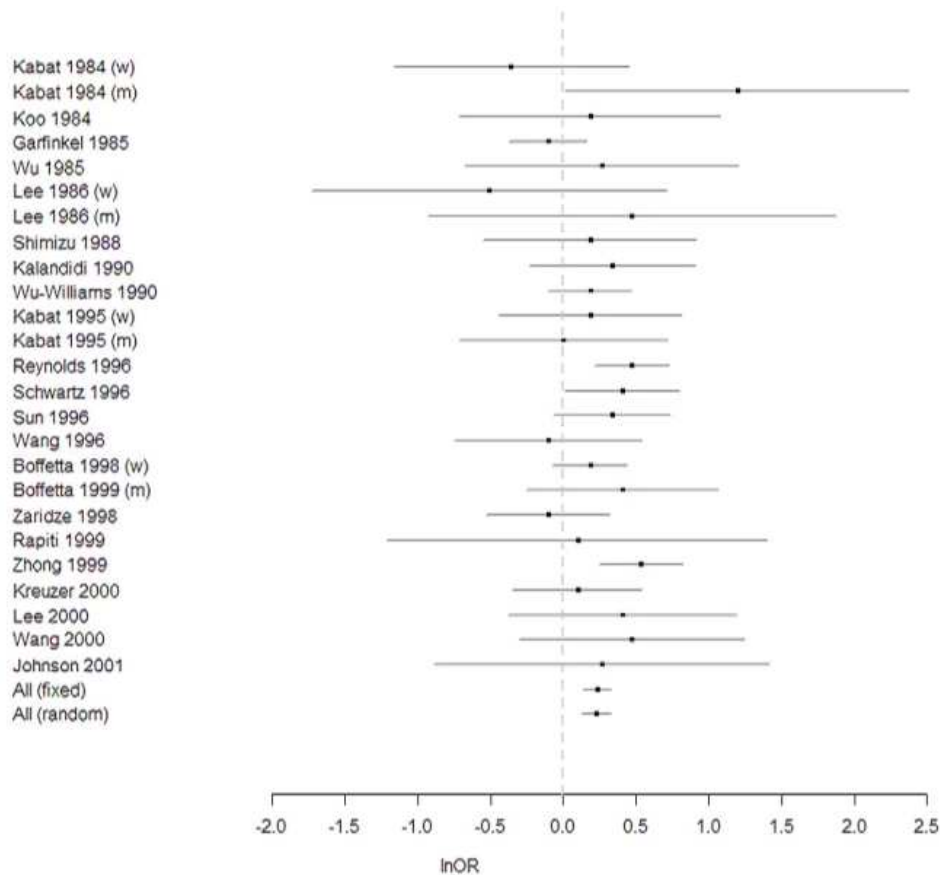


Abbildung 10: Forest Plot der direkten Replikation (Pahl, 2009, S. 75)

4.2.2 Dosage-Response Analysen

An dieser Stelle werden die Berechnung über die Gruppen mit den jeweils höchsten Intensitäten der einzelnen Studien und die Metaregression über die Auswirkungen der Zeit auf das Lungenkrebsrisiko angegeben.

Das folgende Ergebnis bezieht sich auf die Berechnung einer Gesamteffektstärke lediglich für die Gruppen mit den höchsten Passivrauch-Intensitäten am Arbeitsplatz. Für diese Berechnung können 8 Effektstärken aus 7 Studien verwendet werden. Stayner et al. (2007) erhalten dabei als Gesamteffektstärke nach dem Fixed-Effects Modell einen Wert von 2.01 (CI: 1.55; 2.60) und nach dem Random-Effects Modell von 2.01 mit einem (CI: 1.33; 2.60). Diese Werte sind ebenfalls nahezu identisch mit den selbst berechneten Ergebnissen, nach denen die Gesamteffektstärke nach dem Fixed-Effects Modell 2.02 (CI: 1.56; 2.60) und nach dem Random-Effects Modell 1.98 (CI: 1.50; 2.61) beträgt. Auch an dieser Stelle liefert der Q-Test Hinweise, dass man von Homogenität ausgehen kann ($Q = 7.59$, $df = 7$, $p = .37$), die Varianzkomponente der zufälligen Effekte τ^2 beträgt .01297.

Im Falle der Metaregression bilden Stayner et al. (2007) alle verwendeten Datenpunkte aus den 6 Primärstudien in einer Grafik ab und geben folgende Werte an: $\beta = .011$ und $SE = .003$. Für eine Aussetzung von 45 Jahren Passivrauch

am Arbeitsplatz prognostiziert die Regression demnach, laut Stayner et al. (2007), ein Odds Ratio von 1.63 (CI: 1.45; 1.82). Die eigene Berechnung ergibt kein signifikantes Ergebnis mit einem α von 5%, der berechnete t-Wert (1.93) liegt deutlich unter dem kritischen t-Wert von 2.12. Der Wert für das gepoolte b-Gewicht beträgt .009 mit einem dazugehörigen gepollten SE von .005. Das prognostizierte Odds Ratio über 45 Jahre ETS am Arbeitsplatz in Bezug auf das Lungenkrebsrisiko ergibt demnach einen Wert von 1.48 (CI: 0.96; 2.27). Der Q-Test deutet auf Homogenität hin ($Q = 5.27$, $df = 5$, $p = .38$), die Varianzkomponente der zufälligen Effekte τ^2 beträgt nahezu Null. Die Verteilung der Punkte aus den einzelnen Studien findet sich in Abbildung 11 abgetragen. Zusätzlich wird die Regressionsgerade sowie das vorhergesagte Risiko abgebildet, sowohl nach Stayner et al. (2007), als auch nach den eigenen Berechnungen.

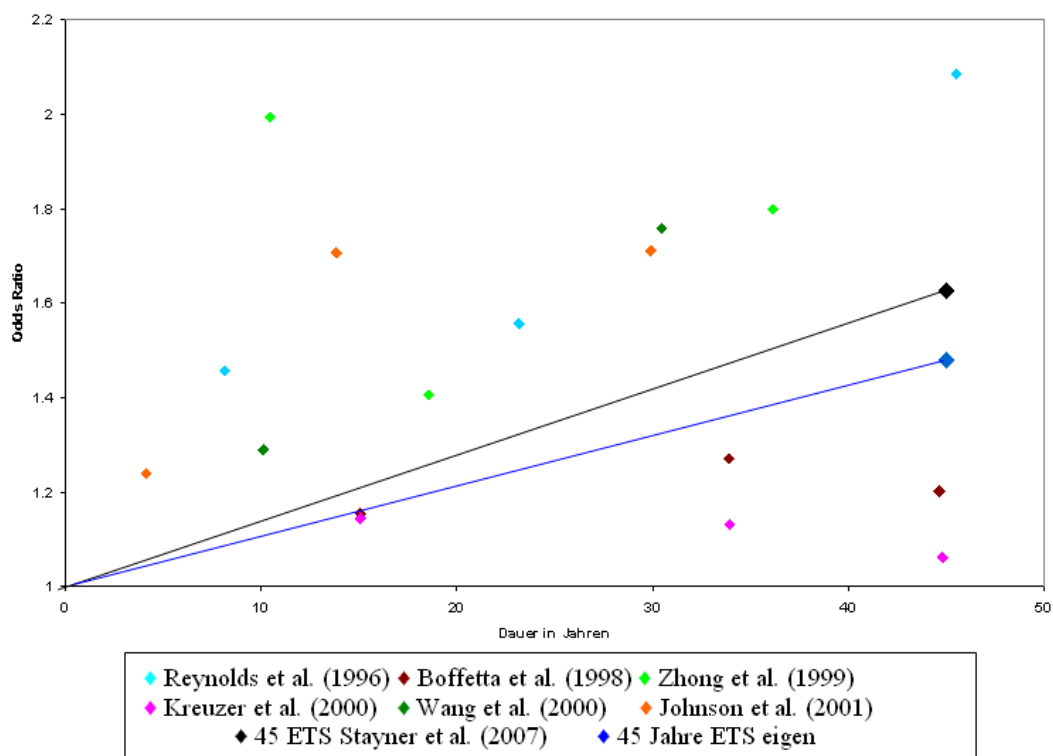


Abbildung 11: Direkte Replikation der Regression zur Vorhersage des Lungenkrebsrisikos über die Dauer der Aussetzung von ETS (Pahl, 2009, S. 77)

4.2.3 Untersuchung auf selektive Veröffentlichung: Trichtergrafikanalyse, N-Q Plots, Rosenberg und Orwin

Der Funnel Plot, den Stayner et al. (2007) präsentieren, wird ebenfalls im Rahmen der Replikation nachgebildet. Abbildung 12 beinhaltet sowohl den Funnel Plot nach Stayner und Kollegen (links), sowie die Nachbildung anhand der direkten Replikation (rechts). Während in beiden Fällen augenscheinlich kein Hinweis auf ein File-Drawer Problem zu finden ist, zeigen sich doch einige Unterschiede zwischen den Plots. Der auffälligste Unterschied liegt in der Kürzung Stayners und seiner Kollegen bei der y-Achse mit den Effektstärken, um die Effektstärke der weiblichen Personen von Kabat (1985) abbilden zu können. Dieser Wert hätte sonst weit außerhalb gelegen, was in der selbst erstellten

Version auch der Fall ist. Bei genauerer Betrachtung fallen noch weitere Unterschiede auf, mehrere der Datenpunkte sind dabei auf der x-Achse (Kehrwert der Varianz der Effektstärke) verschoben. Der Grund für die Verschiebung der Datenpunkte ist dabei wohl in den Rundungsfehlern zu suchen. Zusätzlich spielt es eine große Rolle, anhand welcher Formel die Varianz und somit die Gewichtung abgebildet wird. Würden exakte Werte bei der Berechnung der Varianz zur Verfügung stehen, hätte die Wahl der Formel keinen Einfluss auf das Ergebnis. Da in der Metaanalyse von Stayner et al. (2007) die Werte jedoch auf eine Nachkommastelle gerundet wurden, ergeben sich teilweise große Unterschiede durch die Wahl der Formel. Als Beispiel seien die beiden folgenden Formeln vorgestellt, die jeweils zur Berechnung der Varianz eingesetzt werden könnten:

$$v = \left(\frac{\ln(OG) - \ln(OR)}{1.96} \right)^2 \quad (14)$$

wobei $\ln(OG)$ für den logarithmisierten Wert der oberen Grenze des Konfidenzintervalls steht und

$$v = \frac{[\ln(OG) - \ln(UG)]^2}{(2 * 1.96)^2} \quad (15)$$

wobei $\ln(UG)$ für den logarithmisierten Wert der unteren Grenze des Konfidenzintervalls steht. Aus diesem Grund wurde für die vorliegende Arbeit einheitlich Formel 14 zur Berechnung der Residualvarianz verwendet, sofern adjustierte Werte vorlagen.

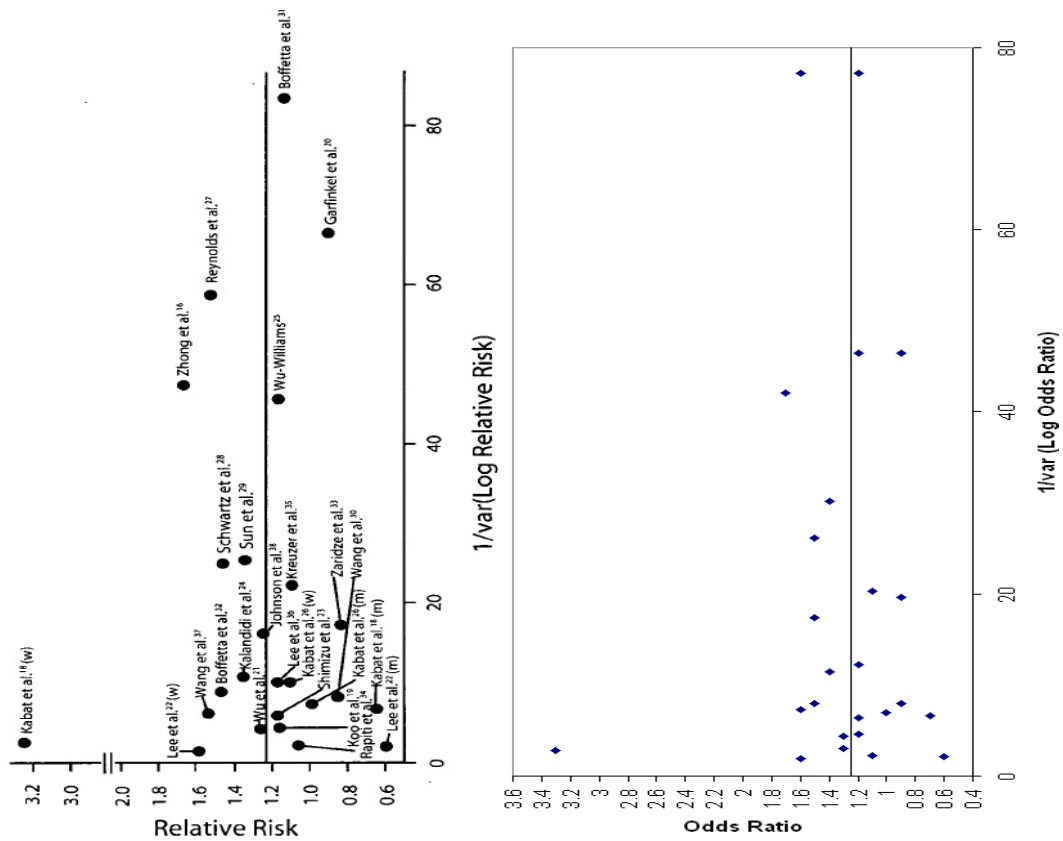


Abbildung 12: Vergleich der Funnel Plots zwischen Original (links) und direkter Replikation (rechts) (Pahl, 2009, S. 80)

Der Normal Quantile Plot über die Effektstärken zur Berechnung der Gesamteffektstärke aller 22 Studien bei Stayner et al. (2007) wird in folgender Abbildung dargestellt.

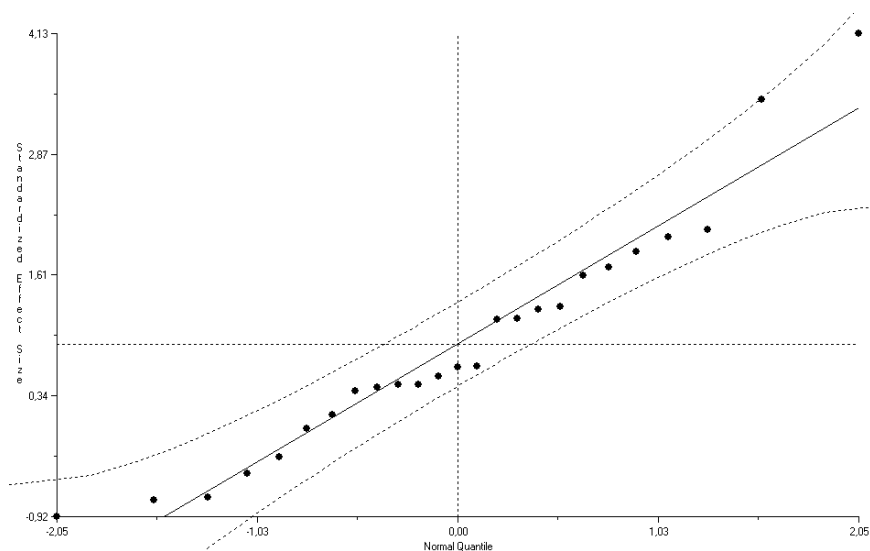


Abbildung 13: Normal Quantile Plot über die Daten der direkten Replikation (Pahl, 2009, S. 92)

Alle Effektstärken liegen innerhalb des 95%-Konfidenzintervalls, wenn auch die Studie von Zhong et al. (1999) beinahe auf der Intervallsgrenze liegt (oben rechts zu sehen). Die Effektstärken sind somit annähernd normalverteilt. Auch der

Normal Quantile Plot lässt auf keinen Publication Bias schließen (Pahl, 2009, S. 92 und Appendix D).

Berechnet man die Ansätze nach Rosenthal (1979), Orwin (1983) und Rosenberg (2005) über die Daten, die zur Berechnung der Gesamteffektstärke in der direkten Replikation anhand des Originalartikels verwendet wurden, so erhält man für die Methode nach Rosenthal einen Wert von 156.5 Studien (bzw. unabhängige Effektstärken), die notwendig wären, um das Ergebnis unterhalb des 5%-Niveaus zu senken. Nach der Methode von Orwin wird ein Wert von 5.8 ausgegeben, was bedeutet, dass noch sechs Studien (bzw. unabhängige Effektstärken) notwendig wären, um den ermittelten Effekt unterhalb das Niveau eines kleinen Effektes zu drücken. Die Methode nach Rosenberg liefert weiterhin unter der Annahme einer Normalverteilung einen Wert von 249.6, was bedeutet, dass 250 Studien (bzw. unabhängige Effektstärken) mit Nulleffekt notwendig wären, um das Ergebnis unter ein Signifikanzniveau von 5% zu senken (Pahl, 2009, S. 91 und Appendix D).

4.2.4 Leave-one-out Analysen

Die Sensitivitätsanalyse in der Arbeit von Stayner et al. (2007) ergab Gesamteffektstärken zwischen 1.18 und 1.27, wobei daraus nicht hervorgeht, ob es sich bei diesen Berechnungen um Ergebnisse nach dem Fixed-Effects oder dem Random-Effects Modell handelt. Nach den eigenen Berechnungen kommt es durch diese Methode zu Gesamteffektstärken zwischen $OR = 1.23$ (CI: 1.11; 1.35) (ohne die Studie von Reynolds et al., 1996) und $OR = 1.33$ (CI: 1.21; 1.46) (ohne die Studie von Garfinkel et al., 1985). In beiden Fällen macht es dabei keinen Unterschied, ob nach dem Fixed-Effects oder nach dem Random-Effects Modell analysiert wird.

Beschränkt sich dieser Ansatz nur auf die sieben Studien, die Angaben zu Personen mit der jeweils höchsten Intensität an Passivrauch am Arbeitsplatz bieten, so liegen die berechneten Werte für die Gesamteffektstärke von Stayner et al. (2007) zwischen $OR = 1.73$ und 2.12, die selbst berechneten Odds Ratios über die betreffenden Studien zwischen 1.74 (CI: 1.29; 2.34) (ohne die Studie von Zhong et al., 1999) und einem Wert von 2.24 (CI: 1.68; 2.97) (ohne die Studie von Kabat et al., 1984). Es macht dabei erneut keinen Unterschied, ob das Fixed-Effects oder das Random-Effects Modell als Grundlage verwendet werden.

Wird der leave-one-out-Ansatz auf die Metaregression angewendet, so erhalten Stayner et al. (2007) sowohl für β als auch für den Standardfehler Werte von .009 bis .014. Nach den eigenen Berechnungen resultieren folgende Ergebnisse: das gepoolte b-Gewicht liegt zwischen .006 (ohne die Studie von Wang et al., 2000) und .011 (jeweils ohne die Studie von Zhong et al., 1999 und ohne die Studie von Kreuzer et al., 2000). Durch die gleiche Gewichtung aller verwendeten Studien liegt der SE-Wert jeweils bei .005.

Eine andere Art der Sensitivitätsberechnung bei Stayner et al. (2007), ist die Bestimmung der Gesamteffektstärke alleine für die adjustierten Studien. Der von Stayner und Kollegen berechnete Wert beträgt dabei $OR = 1.25$ (CI: 1.13; 1.38), der selbst berechnete Wert liegt bei $OR = 1.29$ (CI: 1.18; 1.42) nach dem Fixed-Effects Modell und $OR = 1.29$ (CI: 1.16; 1.43) nach dem Random-Effects Modell.

Auch in diesem Fall bleibt die Homogenität erhalten ($Q = 18.83$, $df = 16$, $p = .28$), die Varianzkomponente der zufälligen Effekte τ^2 beträgt .0070 (Pahl, 2009, S. 77 und Appendix A-C).

4.2.5 Artefaktkorrektur und Variation der Gewichtungsbasis

Eine Korrektur anhand der Daten aus dem Artikel von Stayner et al. (2007) ist nicht möglich, da hier nicht alle benötigten Angaben gegeben sind. In die Gewichtung der Effektstärken muss unter anderem die Gesamtstichprobe der jeweiligen Studien mit eingehen (siehe Formel (7) und (8)). Stayner und Kollegen geben allerdings nur „Case“ Zahlen an, d.h. die Menge an Lungenkrebspatienten, die in die jeweilige Studie mit eingegangen sind; die Größe der Kontrollgruppe fehlt. Da das Verhältnis der Case-Gruppe zu der Kontrollgruppe zum Teil erheblich voneinander abweicht (zwischen 1:1 bis hin zu 1:17 bei Lee et al. (1986)) ist eine Gewichtung über die Case-Zahlen alleine nicht zu empfehlen. Dadurch würden die Primärstudien eine andere Gewichtung erfahren als bei der Originalarbeit und Studien mit unverhältnismäßig großem Kontrollgruppenanteil würden mit einem größeren Gewicht und damit einem größeren Anteil an der Gesamteffektstärke eingehen.

4.3 Replikation des Gesamteffekts anhand der neu codierten Primärstudien

Die Darstellung der Ergebnisse folgt an dieser Stelle ebenfalls der Reihenfolge von Stayner et al. (2007). Zunächst wird die Gesamteffektstärke mit den Moderatoranalysen über die key study variables dargestellt, dann die Subgruppenanalysen der Dosage Response Analysen, gefolgt von den verschiedenen Sensitivitätsanalysen und Untersuchungen zur Überprüfungen eines möglichen Publication Bias.

4.3.1 Schätzung des Gesamteffekts

Nach dem Vorgehen der Effektstärkenauswahl, wie es bei Stayner et al. (2007) beschrieben wird, sind anhand der 22 Primärstudien für die Berechnung der Gesamteffektstärke 27 Effektstärkenschätzer durch die Neucodierung ermittelbar. Die Gesamteffektstärke darüber beträgt $OR = 1.27$ (CI: 1.16; 1.40) nach dem Fixed-Effects Modell und $OR = 1.27$ (CI: 1.15; 1.40) nach dem Random-Effects Modell. Nach dem Ergebnis des Q-Tests kann man auch in diesem Fall von Homogenität der Primärstudien ausgehen ($Q = 27.09$, $df = 26$, $p = .40$), die Varianzkomponente der zufälligen Effekte τ^2 beträgt .0027. Die einzelnen Effektstärken sowie die daraus ermittelte Gesamteffektstärke samt der einzelnen Konfidenzintervalle werden in der folgenden Abbildung in Form eines Forest Plots abgetragen.

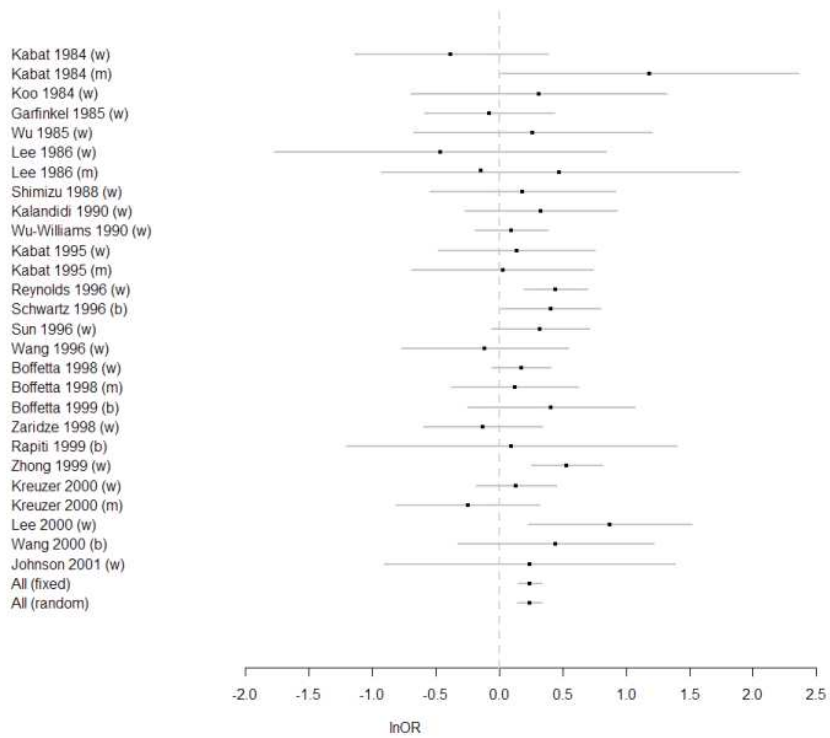


Abbildung 14: Forest Plot anhand der neucodierten Daten (Pahl, 2009, S. 83)

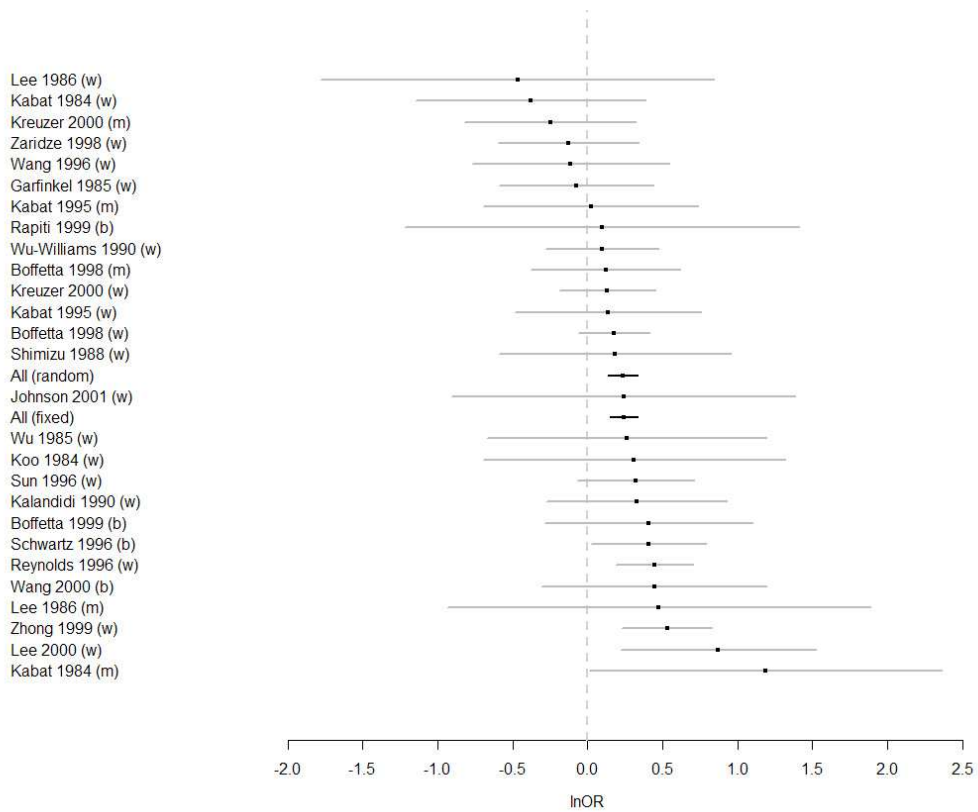


Abbildung 15: Geordneter Forest Plot anhand der neucodierten Daten (Pahl, 2009, S. 83)

Die Moderatoranalysen ergeben anhand der neu codierten Daten nach dem Fixed-Effects Modell mit Moderatoren und dem Mixed-Effects Modell drei signifikante Moderatoren auf dem 95%-Niveau: die Autoren Gruppe um Boffetta (Q = 4.09, p = .04), die statistische Korrektur über Krebs in der Familienhistorie (Q = 4.75, p = .03) und die statistische Korrektur über Anstellungen in risikoreichen Berufen (Q = 4.75, p = .03). Weitere Details finden sich in Pahl (2009, S. 81, Appendix B-a).

4.3.2 Dosage-Response Analysen

An dieser Stelle folgen die Ergebnisse der Subgruppenanalysen über die Intensität und die Dauer von ETS am Arbeitsplatz in Bezug auf die Risiken an Lungenkrebs zu erkranken, durch die Neucodierung der Primärstudien. Zur Berechnung der Gesamteffektstärke mit den Gruppen der höchsten Intensität können durch die Neucodierung insgesamt 11 Effektstärkenschatzer über 7 Studien gewonnen werden. Die Gesamteffektstärke über diese Subgruppen der Studien ergibt nach dem Fixed-Effects Modell einen Wert von 1.92 (CI: 1.51; 2.44) und nach dem Random-Effects Modell einen Gesamtwert von 1.92 (CI: 1.51; 2.44). Unter Verwendung des Q-Tests, zur Ermittlung möglicher Heterogenität zwischen den Effektstärken, kann erneut von empirischer Homogenität gesprochen werden (Q = 10.17, df = 10, p = .43), die Varianzkomponente der zufälligen Effekte τ^2 beträgt .0029. Die verwendeten Effektstärken sind mit weiteren Informationen in Tabelle 3 abgebildet.

Tabelle 3: Effektstärken und 95%-CI für die Berechnung der Gruppen mit den höchsten Intensitäten (Pahl, 2009, S. 84)

Autoren	Geschlecht	Intensitätsmessung	OR	CI-	CI+
Lee et al. (1986)	Männlich	average/a lot	.46	.05	4.65
Lee et al. (1986)	Weiblich	average/a lot	.23	.01	4.10
Kalandidi et al. (1990)	Weiblich	Some	2.17	.83	5.70
Kabat et al. (1995)	Männlich	High	1.21	.47	3.13
Kabat et al. (1995)	Weiblich	High	1.35	.64	2.84
Boffetta et al. (1998)	Männlich	Mehr als 88.9: levels * hours/day * years	2.00	1.03	3.90
Boffetta et al. (1998)	Weiblich	Mehr als 88.9: levels * hours/day * years	1.87	1.10	3.20
Zhong et al. (Zhong)	Weiblich	Mehr als 4 Mitarbeiter, die rauchen	3.00	1.80	4.90
Kreuzer et al. (2000)	Männlich	Mehr als 100600: hours * level of smokiness	1.08	.35	3.33
Kreuzer et al. (2000)	Weiblich	Mehr als 100600: hours * level of smokiness	2.52	1.11	5.69
Johnson et al. (2001)	Weiblich	Größer/gleich 64: Anzahl rauchende Arbeitskollegen * Jahre am Arbeitsplatz	1.58	.60	4.00

Nach der Methode von Greenland und Longnecker (1992) ergibt sich letztendlich ein gepooltes b-Gewicht von .008 mit einem dazugehörigen gepollten SE von .002, was bei einem $\alpha = .05$ statistische Signifikanz bedeutet ($t_{krit} = 2.12 < t_{ber} = 3.50$). Das prognostizierte OR über 45 Jahre ETS am Arbeitsplatz in Bezug auf das Lungenkrebsrisiko, ergibt demnach einen Wert von 1.46 (CI: 1.16; 1.83). Der Q-Test ergibt in diesem Fall ein signifikantes Ergebnis, was auf Heterogenität

schließen lässt ($Q = 15.75$, $df = 5$, $p = .01$), die Varianzkomponente der zufälligen Effekte τ^2 beträgt .0001. Dass das Ergebnis der Metaregression anhand der Neucodierung der Daten für Heterogenität spricht, während die direkte Replikation (wie auch bei Stayner und Kollegen in ihrer Arbeit) Homogenität nachweist, wird vermutlich an der Studie von Kreuzer et al. (2000) liegen, die an dieser Stelle anders abgebildet wird. Die Regression wird in Abbildung 16 grafisch dargestellt.

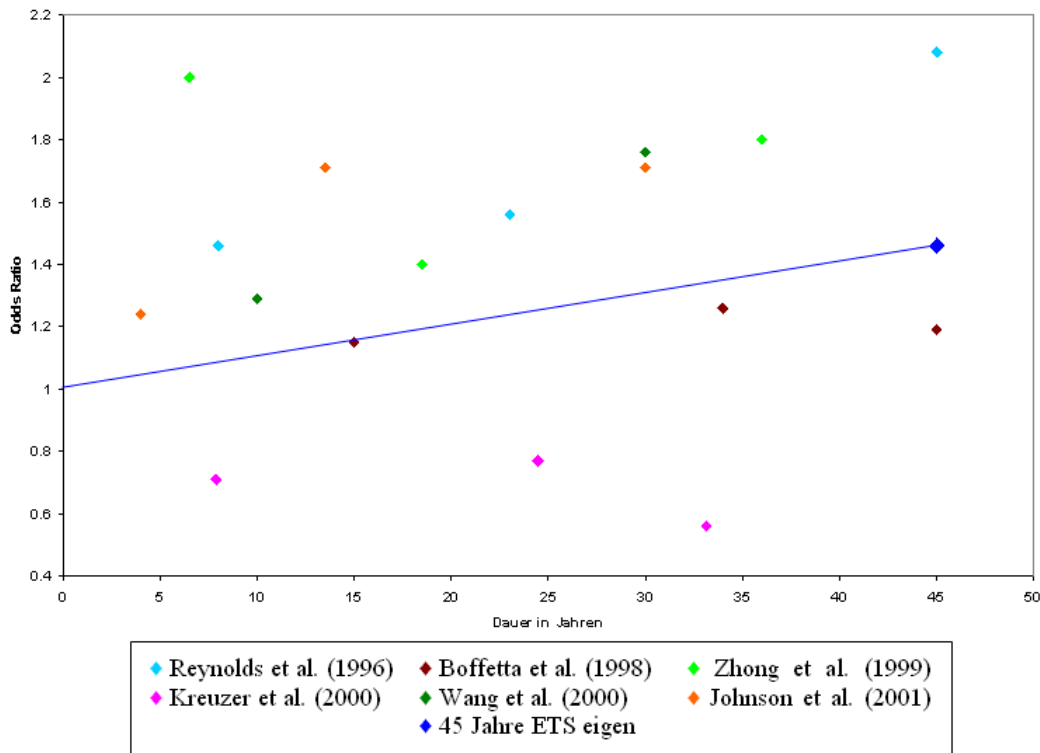


Abbildung 16: Replikation der Regression zur Vorhersage des Lungenkrebsrisikos über die Dauer der Aussetzung von ETS anhand neucodierter Daten (Pahl, 2009, S. 85)

4.3.3 Untersuchung auf selektive Veröffentlichung: Trichtergrafikanalyse, N-Q-Plots, Rosenberg und Orwin

Wird ein Funnel Plot über die 27 Effektstärken zur eigenen Berechnung der Gesamteffektstärke gezeichnet, ergibt sich augenscheinlich erneut kein Hinweis für einen Publication Bias, unabhängig ob der Funnel durch den Kehrwert der Varianz oder den Standardfehler abgetragen wird. Beide Versionen sind in der folgenden Abbildung dargestellt, links mit dem Kehrwert der Varianz und rechts mit SE.

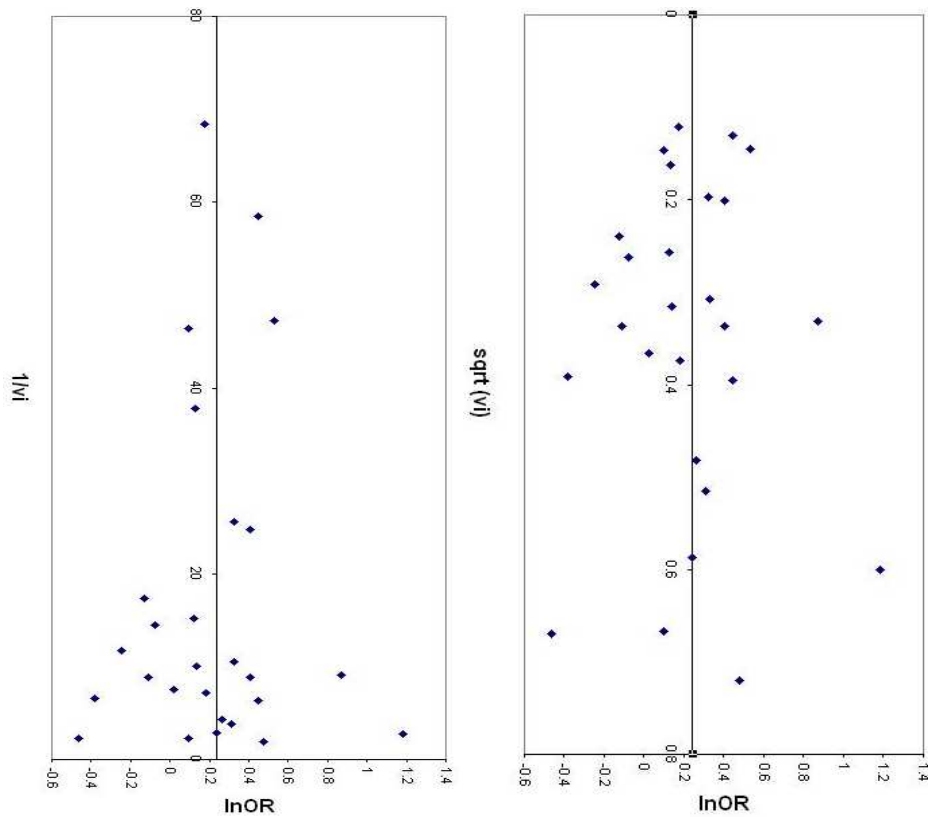


Abbildung 17: Vergleich der Funnel Plots der Neucodierung durch Abbildung mit dem Kehrwert und des Standardfehlers (Pahl, 2009, S. 87 und Appendix B-d)

Die folgende Abbildung zeigt den Normal Quantile Plot über die 27 Effektstärken aus der Neucodierung zur Berechnung der Gesamteffektstärke.

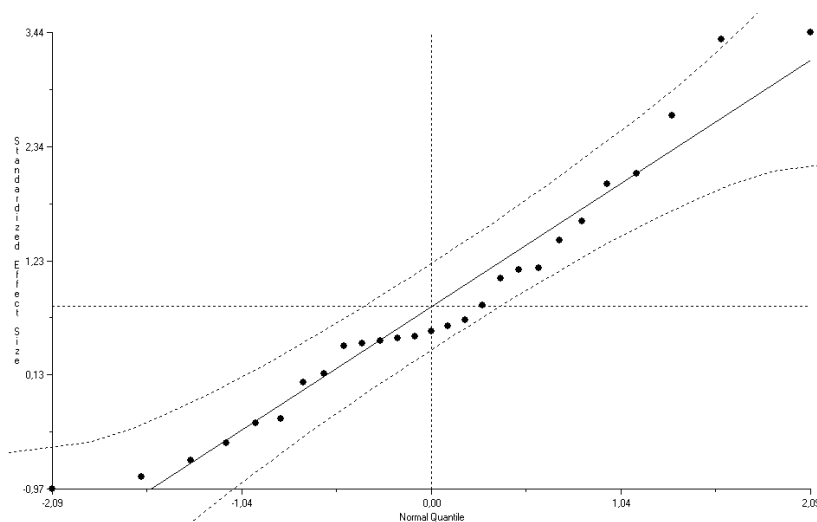


Abbildung 18: Normal Quantile Plot über die Daten der Neucodierung (Pahl, 2009, S. 93)

Die Studie von Reynolds et al. (1996) fällt zwar leicht aus dem Konfidenzintervall heraus (oben rechts zu sehen), da diese Abweichung jedoch sehr gering zu sein scheint, spricht nichts dagegen, die betreffende Studie in den Berechnungen zu belassen. Auch nach dem Normal Quantile Plot kann somit an dieser Stelle kein Publication Bias festgestellt werden (siehe auch Pahl, 2009, S. 93, Appendix D).

Wendet man die drei genannten Methoden nach Rosenthal (1979), Orwin (1983) und Rosenberg (2005) auf die Effektstärken an, die durch die Neucodierung zur Gesamteffektstärkenberechnung gewonnen wurden, erhält man nach Rosenthal (1979) einen Wert von 141.3 und nach Rosenberg (2005) ein Ergebnis von 224.3 Studien (bzw. unabhängige Effektstärken) mit Nulleffekt, die notwendig wären, um das Ergebnis unter ein Signifikanzniveau von 5% zu senken. Nach der Formel von Orwin (1983) resultiert hingegen ein Wert von 5.4, d.h. es wäre auch an dieser Stelle 6 Studien (bzw. unabhängige Effektstärken) notwendig, um die Gesamteffektstärke unter einen kleinen Effekt zu senken (Pahl, 2009, S. 92 und Appendix D).

4.3.4 Leave-one-out Analysen

Die Sensitivitätsanalyse nach dem replizierten Vorgehen von Stayner et al. (2007) ergibt dabei einen Minimalwert von OR = 1.23 (CI: 1.11; 1.36) sowohl nach dem Fixed-Effects Modell, wie nach dem Random-Effects Modell und einen Maximalwert von OR = 1.30 (CI: 1.18; 1.44) erneut für beide Modelle. Die kleinste Gesamteffektstärke wird dabei ohne die Studie von Reynolds et al. (1996), die größte Gesamteffektstärke ohne die Studie von Kreuzer et al. (2000), errechnet.

Wird der leave-one-out-Ansatz über die Studien zur Intensitätsberechnung angewendet, so beträgt der Minimalwert sowohl für das Fixed-Effects Modell, wie auch für das Random-Effects Modell, OR = 1.68 (CI: 1.28; 2.20) (ohne die Studie von Zhong et al., 1999) und der Maximalwert OR = 2.08 (CI: 1.60; 2.70), indem die Berechnung ohne die Studie von Kabat et al. (1995) vorgenommen wird.

Die angepassten und gepoolten Regressionsgewichte über die Auswirkungen der Dauer von Passivrauch am Arbeitsplatz in Bezug auf das Lungenkrebsrisiko, variieren durch das Weglassen einer der Studien zwischen einem Wert von .006 (ohne die Studie von Reynolds et al., 1996) und einem Wert von .011 (ohne die Studie von Kreuzer et al., 2000), der Standardfehler variiert dabei zwischen Werten von .0025 (ohne die Studie von Wang et al., 2000) und .0031 (ohne die Studie von Boffetta et al., 1998). Interessant in diesem Zusammenhang ist, dass die Regression ohne die Studie von Reynolds et al. (1996) bei einem Alpha-Niveau von 5% nicht signifikant wird. Die prognostizierte Effektstärke liegt in diesem Fall bei OR = 1.28 (CI: 0.96; 1.71), während die Regression ohne die Studie von Kreuzer et al. (2000) deutlich signifikant bleibt (OR = 1.67; CI: 1.30; 2.15). In Abbildung 19 sind vier verschiedene Formen der resultierenden Regressionsgerade abgebildet.

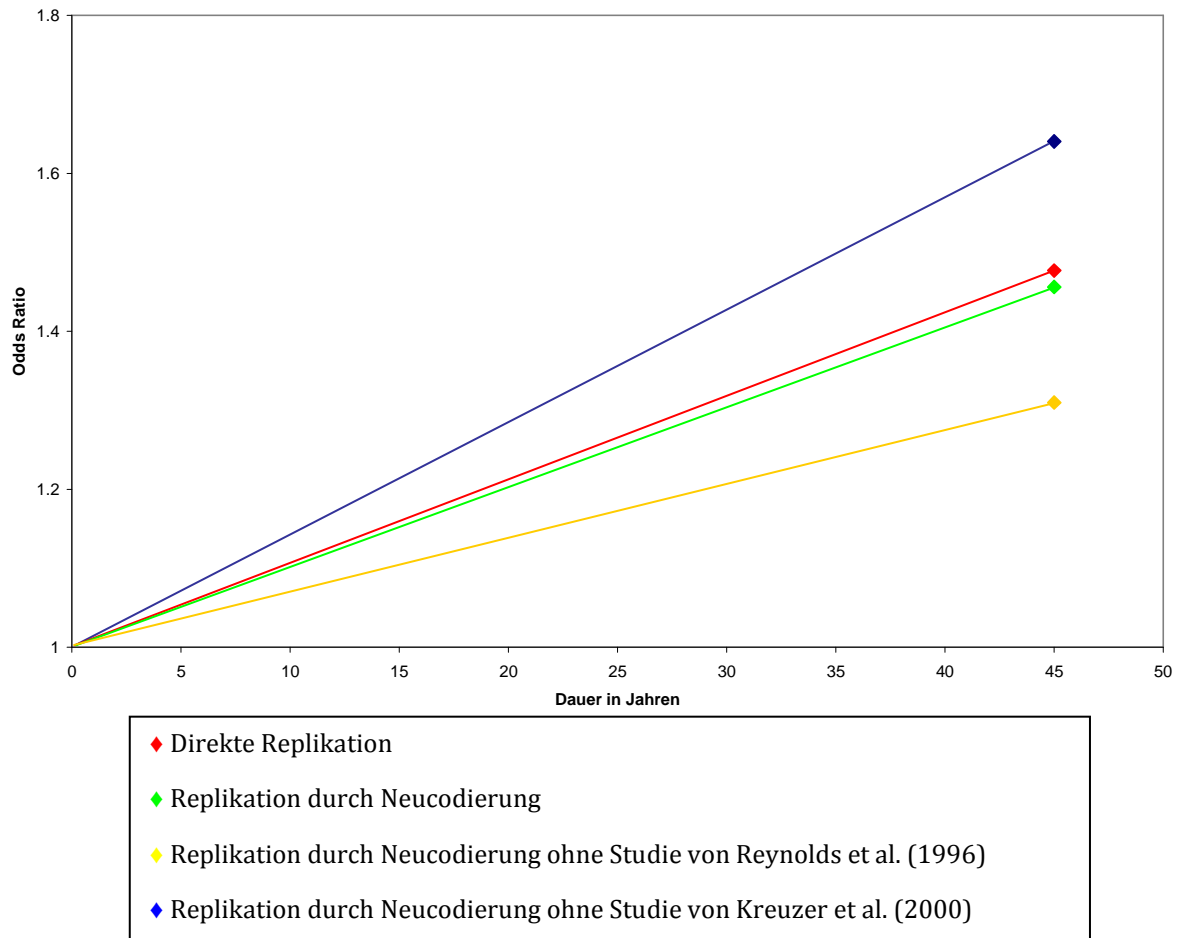


Abbildung 19: Replikation der Regression zur Vorhersage des Lungenkrebsrisikos über die Dauer der Aussetzung von ETS in verschiedenen Varianten

Berechnet man eine Gesamteffektstärke lediglich anhand der statistisch korrigierten Studien, so erhält man dafür nach dem Fixed-Effects Modell einen Wert von $OR = 1.29$ (CI: 1.17; 1.44) und für das Random-Effects Modell einen Wert von $OR = 1.29$ (CI: 1.15; 1.44). Weitere Details finden sich bei Pahl (2009, S. 86 und Appendix B-c).

4.3.5 Artefaktkorrektur und Variation der Gewichtungsbasis

In Tabelle 4 sind die Ergebnisse der Reliabilitätskorrekturen der neu kodierten Primärstudien zu finden. Aus der Studie von Ashton (2000) wurde für die Güte von ärztlichen Urteilen ein Wert von $r_{yy} = .73$ übernommen, aus Pron et al. (1988) eine Kappa von $r_{xx} = .43$ für die Reliabilität von Patientenaussagen zu ETS am Arbeitsplatz. Daraus ergibt sich ein Disattenuationsfaktor von $a = .56$.

Die korrigierte Gesamteffektstärke $OR_c = 1.41$ (1.14; 1.76) ist nominal größer als die unkorrigierte von $OR = 1.24$ (1.09; 1.40). Dieser Effekt ist allerdings nicht signifikant, da sich die Konfidenzintervalle überschneiden. Die Homogenitätsanalyse hat ergeben, dass ein Varianzanteil von 16.7% erklärt wurde, die 75% Regel wurde also deutlich unterschritten, d.h. der Effekt der

Passivrauchexposition auf Lungenkrebs könnte von einer noch unbekanntem Drittvariable beeinflusst sein.

Tabelle 4: Ergebnisse der Artefaktkorrektur mit den Effektstärken, die sich aus der Primärstudienrekodierung ergeben haben. Die unkorrigierten Werte ORo und CIo können direkt mit den korrigierten Werten ORc und CIc verglichen werden.

Studie	ORo*	CIo**	ORc	CIc
Kabat 1984 (w)	0.68	0.53; 0.87	0.51	0.32; 0.79
Kabat 1984 (m)	3.27	2.11; 5.45	11.47	3.92; 1.0151***
Koo 1984 (w)	1.36	1.30; 1.43	1.72	1.57; 1.87
Garfinkel 1985 (w)	0.93	0.83; 1.04	0.88	0.72; 1.06
Lee 1986 (w)	0.63	0.47; 0.83	0.44	0.25; 0.72
Lee 1986 (m)	1.61	1.43; 1.82	2.31	1.86; 2.91
Kalandidi 1990 (w)	1.39	1.31; 1.47	1.77	1.60; 1.96
Wu-Williams 1990 (w)	1.10	1.06; 1.15	1.18	1.10; 1.26
Kabat 1995 (w)	1.15	1.12; 1.17	1.27	1.21; 1.32
Kabat 1995 (m)	1.02	0.95; 1.10	1.04	0.92; 1.17
Reynolds 1996 (w)	1.56	1.40; 1.74	2.19	1.81; 2.67
Schwartz 1996 (b)	1.50	1.37; 1.64	2.04	1.73; 2.40
Wang 1996 (w)	0.89	0.79; 1.02	0.82	0.66; 1.03
Boffetta 1998 (w)	1.19	1.18; 1.20	1.35	1.33; 1.37
Boffetta 1998 (m)	1.13	1.10; 1.16	1.24	1.17; 1.30
Boffetta 1999 (b)	1.50	1.37; 1.64	2.04	1.73; 2.40
Zaridze 1998 (w)	0.88	0.77; 1.01	0.80	0.63; 1.01
Zhong 1999 (w)	1.70	1.47; 1.97	2.56	1.97; 3.39
Kreuzer 2000 (w)	1.14	1.11; 1.17	1.25	1.20; 1.31
Kreuzer 2000 (m)	0.78	0.65; 0.94	0.65	0.47; 0.90
Lee 2000 (w)	2.39	1.79; 3.28	5.12	2.81; 11.54
Johnson 2001 (w)	1.27	1.25; 1.29	1.51	1.46; 1.57
All (random)	1.24	1.09; 1.40	1.40	1.13; 1.73

Werte nach der Vor- und Rückrechnung von OR in Korrelationen, ** CI hier über das Studien N berechnet. Daher wahrscheinlich die Abweichungen vom Stayners Werten, *** Dieser Wert ist noch eine Korrelation. Hier zeigen sich Probleme in der Korrelations/OR Umrechnung

4.4 Auslotung des möglichen Ergebnisspektrums des Gesamteffekts unter Einbezug der im Stellschraubenmodell vorgesehenen subjektiven Entscheidungen

4.4.1 Schätzung des Gesamteffekts auf Basis der Maximalwerte

In Tabelle 5 sind alle gewählten Effektstärken der einzelnen Primärstudien samt den jeweiligen Stellschraubendrehungen abgebildet, indem jeweils die nominal größten Effektstärken aus den einzelnen Primärstudien gezogen wurden. Zusätzlich wird vermerkt, ob es sich bei der gewählten Effektstärke um den Wert einer Subgruppe handelt (z.B. für eine bestimmte Aussetzungsdauer) oder um einen Gesamtwert.

Tabelle 5: Auszug der jeweils maximalen Effektstärken aller Primärstudien, sowie die jeweilige Stellschraubenstellung dazu

Autoren	OR	CI-	CI+	Adjustiert	Geschlecht	Definition	Subgruppe
Kabat (1984)	3.27	1.01	10.62	Nein	männlich	mittel	Nein
Koo (1984)	1.36	.50	3.74	Nein	weiblich	mittel	Nein
Garfinkel (1985)	.93	.73	1.18	Ja	weiblich	mittel	Ja ¹⁾
Wu (1985)	1.30	.50	3.30	Nein	weiblich	mittel	Nein
Lee (1986)	2.76	.64	11.95	Nein	männlich	mittel	Ja ²⁾
Shimizu (1988)	1.20	.60	2.60	Nein	weiblich	mittel	Nein
Kalandidi (1990)	2.17	.82	5.70	Nein	weiblich	mittel	Ja ³⁾
Wu-Williams (1990)	1.22	.95	1.57	Nein	weiblich	mittel	Nein
Kabat (1995)	1.35	.64	2.84	Ja	weiblich	mittel	Ja ⁴⁾
Reynolds (1996)	2.08	1.35	3.20	Ja	weiblich	mittel	Ja ⁵⁾
Schwartz (1996)	1.50	1.00	2.20	Ja	beides	mittel	Nein
Sun (1996)	2.92	1.89	4.49	Ja	weiblich	weit	Nein
Wang (1996)	.89	.46	1.73	Nein	weiblich	mittel	Nein
Boffetta (1998)	2.07	1.33	3.21	Ja	beides	mittel	Ja ⁶⁾
Boffetta (1999)	2.13	.77	5.90	Nein	beides	weit	Ja ⁷⁾
Zaridze (1998)	1.02	.65	1.60	Nein	weiblich	mittel	Nein
Rapiti (1999)	1.10	.30	4.10	Ja	beides	mittel	Nein
Zhong (1999)	3.00	1.80	4.90	Ja	weiblich	mittel	Ja ⁸⁾
Kreuzer (2000)	2.75	1.05	7.24	Nein	weiblich	mittel	Ja ⁹⁾
Lee (2000)	2.612	1.77	3.87	Nein	weiblich	weit	Ja ¹⁰⁾
Wang (2000)	1.76	.50	5.60	Ja	beides	mittel	Ja ¹¹⁾
Johnson (2001)	1.98	.80	4.90	Ja	weiblich	mittel	Ja ¹²⁾
All (fixed)	1.51	1.35	1.68	-	-	-	-
All (random)	1.68	1.37	2.06	-	-	-	-

¹⁾ 25 years ²⁾ little exposure ³⁾ some exposure ⁴⁾ high job years ⁵⁾ >30 years ⁶⁾ >89 years x level ⁷⁾ >201 hours/day ⁸⁾ >4 smoking co-workers ⁹⁾ >61000 hours ¹⁰⁾ >40 years ¹¹⁾ >20 Jahre ¹²⁾ 26-64 years

Werden jeweils die größten Effektstärkenbeträge der 22 Primärstudien zur Berechnung einer Gesamteffektstärke herangezogen, ergibt dies nach dem Fixed-Effects Modell ein Odds Ratio über die Studien von 1.51 (CI: 1.35; 1.68) und eine Gesamteffektstärke von OR = 1.68 (CI: 1.37; 2.06) nach dem Random-Effects Modell. Allerdings muss an dieser Stelle erwähnt werden, dass der Homogenitätstest $p = .000$ ausgibt, was bedeutet, dass das Ergebnis heterogen und somit im eigentlichen Sinne nicht interpretierbar ist. Auf eine Subgruppenanalyse oder eine Moderatorenberechnung wird an dieser Stelle jedoch verzichtet, da der nominal größtmögliche Wert einzig zur Darstellung der Maximalgrenze dargestellt wird und keine inhaltliche Aussage enthält (siehe auch Pahl, 2009, S. 89 und Appendix C).

4.4.2 Schätzung des Gesamteffekts auf Basis der Minimalwerte

Wie unter 4.4.1 wurde theorieunabhängig aus jeder Studie der Originaluntersuchung Stayner et al. (2007) eine Effektstärke übernommen,

diesmal die kleinste. Eine Übersicht ist in Tabelle 6 zu finden. Die Stellschraubenausrichtung weicht hierbei von der Ausrichtung unter der „Maximum“ Bedingung stark ab. Gleiches gilt für die gewählten Subgruppen.

Tabelle 6: Auszug der jeweils minimalen Effektstärken aller Primärstudien, sowie die jeweilige Stellschraubenstellung dazu

Autoren	OR	CI-	CI+	Adjustiert	Geschlecht	Definition	Subgruppe
Kabat (1984)	0.68	0.31	1.47	Nein	weiblich	mittel	Nein
Koo (1984)	0.91	0.15	5.37	Nein	weiblich	eng	Nein
Garfinkel (1985)	0.88	0.66	1.18	Nein	weiblich	mittel	Ja ¹⁾
Wu (1985)	1.30	0.50	3.30	Nein	weiblich	NA	Nein
Lee (1986)	0.19	0.023	1.65	Ja	beides	mittel	Nein
Shimizu (1988)	1.20	0.60	2.60	Nein	weiblich	mittel	Nein
Kalandidi (1990)	1.28	0.68	2.37	Nein	weiblich	NA	Ja ²⁾
Wu-Williams (1990)	1.10	0.90	1.60	Ja	weiblich	mittel	Nein
Kabat (1995)	0.94	0.42	2.10	Ja	weiblich	mittel	Ja ³⁾
Reynolds (1996)	1.30	1.01	1.67	Nein	weiblich	mittel	Ja ⁴⁾
Schwartz (1996)	1.35	0.96	1.89	Nein	beides	mittel	Nein
Sun (1996)	1.38	0.94	2.04	Ja	weiblich	weit	Nein
Wang (1996)	0.89	0.45	1.77	Nein	weiblich	mittel	Nein
Boffetta (1998)	0.87	0.63	1.21	Ja	weiblich	weit	Ja ⁵⁾
Boffetta (1999)	0.50	0.20	1.50	Ja	beides	weit	Ja ⁶⁾
Zaridze (1998)	0.88	0.55	1.41	Ja	weiblich	mittel	Nein
Rapiti (1999)	1.10	0.30	4.10	Ja	beides	mittel	Nein
Zhong (1999)	0.81	0.50	1.30	Nein	weiblich	mittel	Ja ⁷⁾
Kreuzer (2000)	0.54	0.278	1.04	Nein	beides	mittel	Ja ⁸⁾
Lee (2000)	1.20	0.50	2.40	Ja	weiblich	mittel	Nein
Wang (2000)	1.29	0.50	3.30	Ja	beides	mittel	Ja ⁹⁾
Johnson (2001)	0.79	0.33	1.88	Ja	weiblich	weit	Ja ¹⁰⁾
All (fixed)	1.04	0.94	1.16	-	-	-	-
All (random)	1.03	0.92	1.16	-	-	-	-

¹⁾ 5 years, ²⁾ minimal, ³⁾ intermediate, ⁴⁾ 1-15 years, ⁵⁾ 0-165 hours/day * year, ⁶⁾ 15+ years since last exposure, ⁷⁾ 1-2h exposure per day, ⁸⁾ >56.200-100.600 hours x level of smokiness, ⁹⁾ >20 y exposure, ¹⁰⁾ Residential+occupational smoker years 1-36

Unter dem Fixed-Effecs Model ergibt sich mit den 22 Studien ein mittleres OR von 1.04 (CI: 0.94; 1.16), unter dem Random-Effecs Modell ein mittleres OR von 1.03 (CI: 0.92; 1.16). Unter der Minimalbedingung ergibt sich kein signifikanter Homogenitätstest (p= .35).

4.4.3 Zusammenfassung beider Ansätze und ausgewählte Befunde aus Pahl (2009) und Würfel (2009)

Anhand der unter 4.4.1 und 4.4.2 präsentierten Ergebnissen erkennt man eindrucksvoll, zu welchen unterschiedlichen Ergebnissen eine Metaanalyse kommen kann, trotzdem dass die gleichen Primärstudien verwendet werden. Die verschiedenen Stellschraubendrehungen machen es möglich.

Das Stellschraubenmodell bietet auch die Möglichkeit einzelne Stellschrauben zu fixieren und andere variabel zu gestalten. Dies könnte durchaus Entscheidungsprozesse konkreter Forschungsarbeiten widerspiegeln, in denen einige Entscheidungen bewusst festgelegt werden, andere aber impliziter Natur sind und sich im Laufe der Untersuchung auch noch ändern könnten. Um die möglichen Unterschiede der zu berechnenden Gesamteffektstärke anhand der vorliegenden Daten nachzuweisen werden vier verschiedene Theorien abgebildet, wie sie bei einer Untersuchung zum Thema Auswirkungen von Passivrauch am Arbeitsplatz entstehen könnten.

Die erste Theorie versucht ein Ergebnis zu ermitteln, dass lediglich für Frauen gilt, die allesamt Passivrauch auf der Arbeit ausgesetzt sind. Die Auswahl der Effektstärken erfolgt dabei nach dem Vorgehen:

- Nur Effektstärken für Frauen werden beachtet, darüber hinaus werden keine Subgruppendaten (z.B. über Intensität) aufgenommen.
- Es wird ausschließlich die mittlere Definition verwendet.
- Die Stellschraube, ob adjustierte oder nicht angepasste Effektstärkenschätzer aufgenommen werden, wird innerhalb von zwei Berechnungen variiert.

Somit werden zwei Gesamteffektstärken berechnet, eine nur über adjustierte Werte und eine lediglich über unadjustierte Effektstärken.

Die zweite Theorie folgt dem gleichen Aufbau mit dem Unterschied, dass an dieser Stelle die Werte für Männer statt der Frauen gewählt werden. Ansonsten ist das Vorgehen identisch mit der ersten Theorie. Auch hier werden zwei Gesamteffektstärken berechnet, einmal über angepasste Werte und einmal für Werte ohne Anpassungen.

Die dritte Theorie legt ein anderes Vorgehen fest. Demnach wird ein Kriterium gewählt, das sich streng an dem Forschungsobjekt von Stayner et al. (2007) richtet, nach dem die Stichprobe der Studien aus Versuchspersonen bestehen sollte, die ausschließlich am Arbeitsplatz ETS ausgesetzt sind. Dies bedeutet konkret, dass lediglich Werte nach der engen Definition aufgenommen werden dürfen. Die anderen beiden Stellschrauben sind nicht starr. Das Vorgehen der dritten Theorie ist demnach folgendermaßen aufgebaut:

- Lediglich Effektstärken der engen Definition werden ausgewählt.
- Wann immer es möglich ist, werden die Effektstärken für Männer und Frauen getrennt aufgenommen, falls dies nicht möglich wird eine Effektstärke für beide Geschlechter gemeinsam und sofern das nicht abgebildet wird eine Effektstärke nur für Frauen gewählt. Andere Subgruppendaten werden nicht verwendet.
- Adjustierte Werte werden unadjustierten Werten vorgezogen.

Die vierte Theorie schließlich ist im Vorgehen nahezu identisch mit der dritten genannten Theorie, mit der Ausnahme, dass statt der engen hier die weite Definition interessiert, also eine Verallgemeinerung auf eine Population, die

überall Passivrauch ausgesetzt sein kann. Das restliche Vorgehen erfolgt analog zu Theorie 3.

Nachdem alle vier beschriebenen Theorien berechnet wurden, zeigt sich folgendes Ergebnis: Die Theorie, die lediglich der engen Definition folgt (Theorie 3), schätzt die größte Gesamteffektstärke über die vier dabei verwendbaren Effektstärken. Demnach ist die Gesamteffektstärke über beide Modelle gleich 1.45 (CI: 1.05; 1.99). Danach folgt die Gesamteffektstärke über die neun verwendbaren Effektstärken zur weiten Definition (Theorie 4) mit einer Gesamteffektstärke von 1.40 (CI: 1.23; 1.61) nach dem Fixed-Effects Modell und 1.44 (CI: 1.14; 1.82) nach dem Random-Effects Modell. Die Testung der ersten Theorie, jeweils die Werte der Frauen nach der mittleren Definition, hat ergeben, dass die Gesamteffektstärke größer ausfällt, wenn die 14 unadjustierten Effektstärken zur Berechnung verwendet werden, als mit den 7 angepassten ORs. Demnach beträgt die Gesamteffektstärke der unadjustierten Daten 1.33 (CI: 1.20; 1.47) nach dem Fixed-Effects Modell und 1.28 (CI: 1.09; 1.51) nach dem Random-Effects Modell gegenüber einer Gesamteffektstärke von 1.23 (CI: 1.08; 1.40) nach beiden Modellen anhand der adjustierten Werte. Die Untersuchung der zweiten Theorie hat folglich die kleinsten Effektgrößen ergeben. Für die Berechnung über die unangepassten Effektgrößen berechnete sich eine Gesamtgröße von 1.11 (CI: 0.83; 1.48) nach dem Fixed-Effects Modell und ein Wert von 1.13 (CI: 0.79; 1.61) nach dem Random-Effects Modell. Wird die Gesamteffektstärke über die adjustierten Werte der Männer berechnet ergibt dies nach dem Fixed-Effects Modell einen Wert von 0.87 (CI: 0.66; 1.14) und nach dem Random-Effects Modell einen Wert von 0.89 (CI: 0.63; 1.27). Allerdings muss zu den Ergebnissen der zweiten Theorie angemerkt werden, dass sowohl zur Berechnung der Gesamteffektstärken mit den unadjustierten Werten, als auch mit den adjustierten Effektgrößen, lediglich die Studien von Boffetta et al. (1998) und Kreuzer et al. (2000) zur Verfügung stehen. Somit werden die Gesamteffektgrößen in diesen Fällen jeweils an zwei Effektstärken berechnet, wobei sich die Versuchspersonen beider Studien, nach Aussage von Stayner et al. (2007), teilweise überschneiden. Die erhaltenen Resultate sind dementsprechend mit Vorsicht zu betrachten.

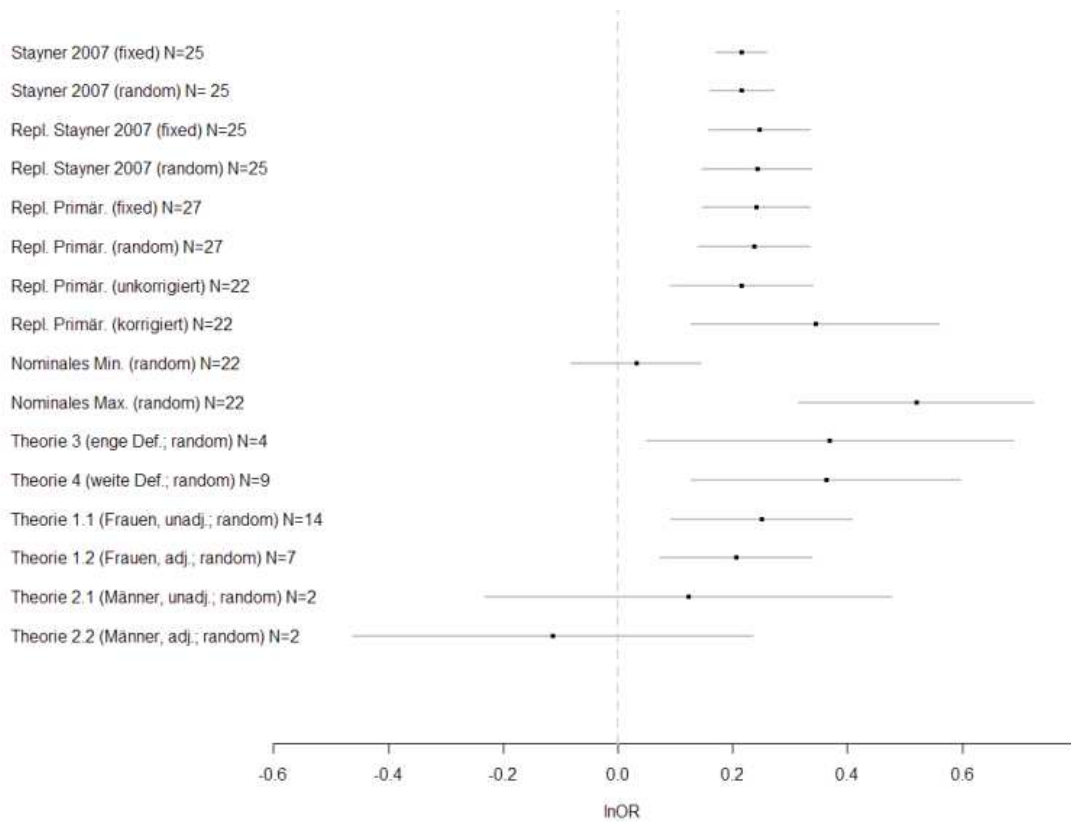


Abbildung 20: Forest Plot aller berechneten Gesamteffektstärken (Pahl, 2009, S. 103).

5 Zusammenfassung und Bewertung

Das übergeordnete Ziel der Studie bestand darin, die *Zuverlässigkeit, Robustheit und Widerlegungssicherheit* der meta-analytischen Befunde bei Stayner et al. (2007) mittels moderner meta-analytischer Verfahren und Techniken zu überprüfen. Die vorliegende Replikation und Erweiterung der Meta-Analyse von Stayner et al. (2007) liefert nun erste Antworten auf die folgenden drei Fragen:

Von welcher Qualität sind die berücksichtigten Primärstudien?

Anhand der zugrunde gelegten Qualitätskriterien muss festgestellt werden, dass die einbezogenen Primärstudien großteils von 'mittlerer' Qualität sind. Es ist zu befürchten, dass deshalb auch die Qualität der meta-analytischen Befunde im Sinne einer 'garbage-in, garbage-out'-Problematik als nicht sehr hoch einzuschätzen sind. Als bedenklich kann der Befund gesehen werden, dass qualitativ hochwertigere Studien mit einer tendenziell geringeren Effektstärkenschätzung systematisch verknüpft zu sein scheinen, d.h. 'bessere' Studien suggerieren einen 'geringeren' Zusammenhang zwischen der Passivrauchexposition und dem Risiko, an Lungenkrebs zu erkranken.

Wie zuverlässig und robust sind die ermittelten Gesamteffekte?

Unter vergleichbaren, d.h. hier *inkonsistent angewandten Codierregeln* wie bei Stayner et al. (2007), sind die Befunde replizierbar. Inkonsistenzen ergaben sich vor allem bei einzelnen codierten Effektstärken, stellenweise wurde auch scheinbar die 'falsche' Effektstärke entnommen. Das sogenannte 'statistische Fruchtsalatproblem' (Brüderl, 2002), d.h. dass Schätzungen für unterschiedliche Populationsparameter vermischt wurden, reflektiert eine weitere Inkonsistenz im Rahmen der Codierung von Primärstudien. Da die Autoren ihr Vorgehen nicht eindeutig nachvollziehbar beschrieben und vor allem die Originaldaten nicht freigegeben haben, bleiben die Gründe für diese Unstimmigkeiten im Dunkeln.

Unter Berücksichtigung und Modellierung der getroffenen subjektiven Entscheidungen während des meta-analytischen Prozesses kann allerdings kaum von robusten Befunden gesprochen werden. Nulleffekte, d.h. die Möglichkeit dass kein Zusammenhang zwischen Passivrauchexposition am Arbeitsplatz und dem Lungenkrebsrisiko besteht, sind im Rahmen des zugrunde gelegten Stellschraubenmodells möglich. Andererseits wäre auf Basis des Stillschraubenmodells ebenfalls denkbar, dass der 'wahre Effekt' bei einer rund 70%-igen Risikosteigerung liegt. Diese verhältnismäßig große Bandbreite an möglichen Ergebnissen verdeutlicht, dass die Primärstudienbasis alles andere als ausreichend erscheint, um eine zuverlässige Aussage über den 'wahren Wert' zu treffen. Abbildung 21 verdeutlicht die mögliche Bandbreite der Befunde unter Berücksichtigung der von Stayner et al. (2007) einbezogenen Studienbasis.

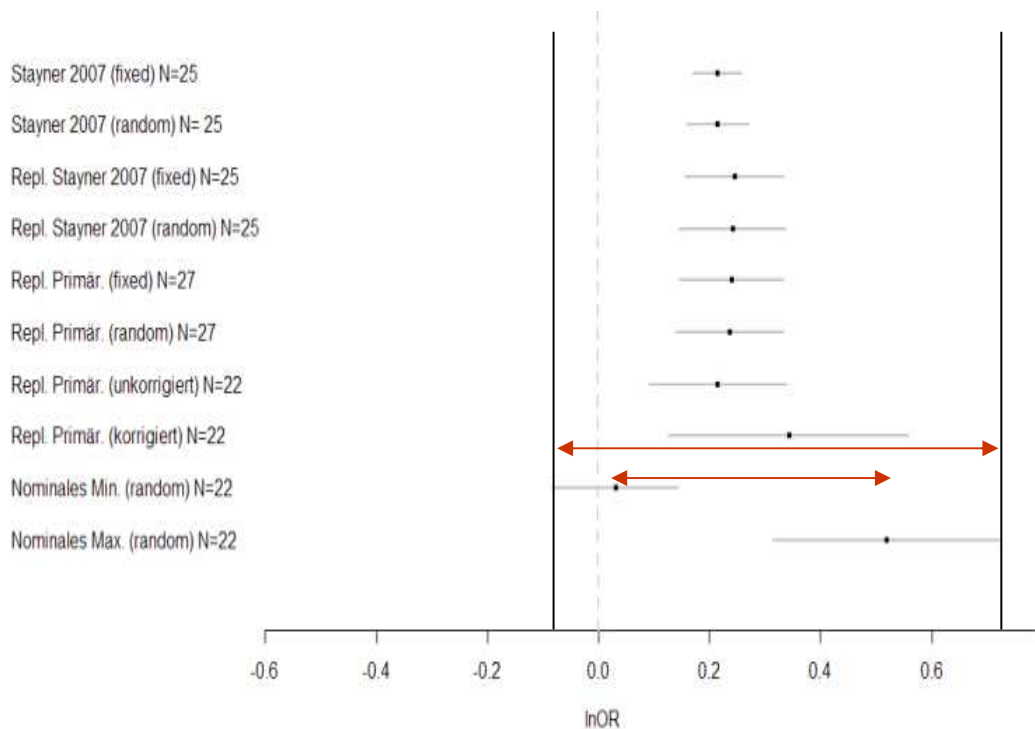


Abbildung 21: Bandbreite der Befunde

Sind die Dosage-Response-Analyse zum Zusammenhang zwischen der Dauer der Passivrauchexposition in Jahren sowie dem Lungenkrebsrisiko robust?

Hier konnte eindeutig gezeigt werden, dass die ermittelten Dosage-Response-Effekte stark voneinander abweichen sobald einzelne Studien ausgeschlossen werden. Bereits der Ausschluss einer einzigen Studie führt dazu, dass der lineare Zusammenhang nivelliert wird, d.h. dass kein statistisch bedeutsamer Zusammenhang zwischen der Dauer der Passivrauchexposition in Jahren sowie dem Lungenkrebsrisiko mehr ermittelbar ist. Andererseits führt der Ausschluss einer anderen Studie zu einem dramatischen Anstieg des Zusammenhangs. Es zeigt sich somit ein ähnliches Bild wie beim geschätzten Gesamteffekt: Die berücksichtigte Primärstudienbasis scheint nicht auszureichen, um eine zuverlässige Aussage über den 'wahren Zusammenhang' zwischen Passivrauchexposition und dem Risiko, an Lungenkrebs zu erkranken, treffen zu können.

In der Gesamtwürdigung der ermittelten Ergebnisse der vorliegenden Replikation und Erweiterung, und unter Berücksichtigung der übergeordneten Fragestellung muss man davon ausgehen, dass die bei Stayner et al. (2007) wiedergegebenen Befunde kaum als zuverlässig, robust und widerlegungssicher einzustufen sind.

6 Literatur

Studien, die in die Metaanalyse einbezogen wurden, sind mit einem Sternchen (*) gekennzeichnet

Ashton, R. H. (2000). A Review and Analysis of Research on the Test±Retest Reliability of Professional Judgment. *Journal of Behavioral Decision Making*, 13, 277-294.

*Boffetta, P., Agudo, A., Ahrens, W., Benhamou, E., Benhamou, S., Darby, S. C., et al. (1998). Multicenter case-control study of exposure to environmental tobacco smoke and lung cancer in Europe. *Journal of the National Cancer Institute*, 90, 1440-1450.

*Boffetta, P., Ahrens, W., Nyberg, F., Mukeria, A., Brüske-Hohlfeld, I., Fortes, C., et al. (1999). Exposure to environmental tobacco smoke and risk of adenocarcinoma of the lung. *International Journal of Cancer*, 83, 635-639.

Bosnjak, M. (2007, Oktober). *Die Methode der Meta-Analyse zur Evidenzbasierung von Gesundheitsrisiken: Mögliche Beiträge der Psychologie*. Eingeladener Vortrag, gehalten auf dem internationalen Symposium zum Thema "Tabakrauch am Arbeitsplatz: Gesundheitsrisiken und Kausalität", 23.-24. Oktober 2007, an der Universität Mannheim.
http://docs.google.com/Presentation?id=ajktzsnh86vs_689c4vtt3

Bosnjak, M., & Viechtbauer, W. (in Druck). Die Methode der Meta-Analyse zur Evidenzbasierung von Gesundheitsrisiken: Beiträge der Sozial-, Verhaltens- und Wirtschaftswissenschaften.

Brüderl, J. (2004). Meta-Analyse in der Soziologie: Bilanz der deutschen Scheidungsursachenforschung oder "statistischer Fruchtsalat"? *Zeitschrift für Soziologie*, 33, 84-86.

Frantz, J. (2000). G3data 1.5.1. From <http://www.frantz.fi/software/g3data.php>.

*Garfinkel, L., Auerbach, O., & Joubert, L. (1985). Involuntary smoking and lung cancer: a case-control study. *Journal of the National Cancer Institute*, 75, 463-469.

Greenland, S., & Longnecker, M. P. (1992). Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *American Journal of Epidemiology*, 135, 1301-1309.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1, 77-89.

Huang, D., Guan, P., Shi, H., He, Q. & Zhou, B. (2008). Reliability and accuracy of interview data in non-smoking female lung cancer case-control study. *Journal of Experimental and Clinical Cancer Research*, 27: 43.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2. Auflage). Thousand Oaks, CA: Sage Publications.

*Johnson, K. C., Hu, J., & Mao, Y. (2001). Lifetime residential and workplace exposure to environmental tobacco smoke and lung cancer in never-smoking women, Canada. *International Journal of Cancer*, 93, 902-906.

*Kabat, G. C., & Wynder, E. L. (1984). Lung cancer in nonsmokers. *Cancer*, 53, 1214-1221.

*Kabat, G. C., Stellman, S. D., & Wynder, E. L. (1995). Relation between exposure to environmental tobacco smoke and lung cancer in lifetime nonsmokers. *American Journal of Epidemiology*, 142, 141-148.

*Kalandidi, A., Katsouyanni, K., Voropoulou, N., Bastas, G., Saracci, R., & Trichopoulos, D. (1990). Passive smoking and diet in the etiology of lung cancer among non-smokers. *Cancer Causes and Control*, 1, 15-21.

*Koo, L. C., Ho, J. H. C., & Saw, D. (1984). Is passive smoking an added risk factor for lung cancer in Chinese women? *Journal of Experimental & Clinical Cancer Research*, 3, 277-284.

*Kreuzer, M., Krauss, M., Kreienbrock, L., Jockel, K. H., & Wichmann, H. E. (2000). Environmental tobacco smoke and lung cancer: a case-control study in Germany. *American Journal of Epidemiology*, 151, 241-250.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

*Lee, C. H., Ko, Y. C., Goggins, W., Huang, J. J., Huang, M. S., Kao, E. L., et al. Lifetime environmental exposure to tobacco smoke and primary lung cancer of non-smoking Taiwanese women. *International Journal of Epidemiology*, 29, 224-231.

*Lee, P. N., Chamberlain, J., & Alderson, M. R. (1986). Relationship of passive smoking to risk of lung cancer and other smoking-associated diseases. *British Journal of Cancer*, 54, 97-105.

Morton, S. C., Adams, J. L., Suttrop, M. J., & Shekelle, P. G. (2004). *Meta-regression approaches: what, why, when, and how? Technical Review 8* (Prepared by Southern California–RAND Evidence-based Practice Center, under Contract No 290-97-0001). AHRQ Publication No. 04-0033. Rockville, MD: Agency for Healthcare Research and Quality.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8, 157-159.

Pahl, H. (2009). *Reanalyse und Erweiterung einer Metaanalyse: Kennzeichnung der Subjektivität durch den Stellschraubenansatz*. Diplomarbeit, Universität Mannheim.

Pron, G. E., Burch, J. D., Howe, G. R. & Miller, A. B. (1988) The Reliability of Passive Smoking Histories Reported in a Case-Control Study of Lung Cancer. *American Journal of Epidemiology*, 127, 267-273.

*Rapiti, E., Jindal, S. K., Gupta, D., & Boffetta, P. (1999). Passive smoking and lung cancer in Chandigarh, India. *Lung Cancer*, 23, 183-189.

*Reynolds, P., von Behren, J., Fontham, E. T., Correa, P., Wu, A., Buffler, P. A. & Greenberg, R. S. (1996). Occupational exposure to environmental tobacco smoke. *The Journal of the American Medical Association*, 275, 441-442.

Rosenberg, M. (2005). The file drawer-problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, 59, 464-468.

Rosenthal, R. (1979). The "File drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe & Huber.

*Schwartz, A. G., Yang, P., & Swanson, G. M. (1996). Familial risk of lung cancer among nonsmokers and their relatives. *American Journal of Epidemiology*, 144, 554-562.

*Shimizu, H., Morishita, M., Mizuno, K., Masuda, T., Ogura, Y., Santo, M., et al. (1988). A casecontrol study of lung cancer in nonsmoking women. *The Tohoku Journal of Experimental Medicine*, 154, 389-397.

Stayner, L., Bena, J., Sasco, A. J., Smith, R., Steenland, K., Kreuzer, M., & Straif, K. (2007). Lung cancer risk and workplace exposure to environmental tobacco smoke. *American Journal of Public Health*, 97, 545-551.

*Sun, X.-W., Dai, X.-D., Lin, C.-Y., Shi, Y.-B., Ma, Y.-Y., & Li, W. (1996). Passive smoking and lung cancer among nonsmoking women in Harbin, China. *International symposium on lifestyle factors and human lung cancer, China 1994. Lung Cancer*, 14: S237.

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gotzsche, P. C., Mulrow, C. D., Pocock, S. J., et al. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) - explanation and elaboration. *Epidemiology*, 18, 805-835.

*Wang, L., Lubin, J. H., Zhang, S. R., Metayer, C., Xia, Y., Brenner, A., et al. (2000). Lung cancer and environmental tobacco smoke in a non-industrial area of China. *International Journal of Cancer*, 88, 139-154.

*Wang, T.-J., Zhou, B.-S., & Shi, J.-P. (1996). Lung cancer in nonsmoking Chinese women: a case-control study. International symposium on lifestyle factors and human lung cancer, China 1996. *Lung Cancer*, 14: S93-S98.

*Wu, A. H., Henderson, B. E., Pike, M. C., & Yu, M. C. (1985). Smoking and other risk factors for lung cancer in women. *Journal of the National Cancer Institute*, 74, 747-751.

*Wu-Williams, A. H., Dai, X. D., Blot, W., Xu, Z. Y., Sun, X. W., Xiao, H. P., et al. (1990). Lung cancer among women in north-east China. *British Journal of Cancer*, 62, 982-987.

Wittmann, W. W., & Matt, G. E. (1986). Meta-Analyse als Integration von Forschungsergebnissen am Beispiel deutschsprachiger Arbeiten zur Effektivität von Psychotherapie. *Psychologische Rundschau*, 37, 20-40.

Würfel, R. S. (2009). Meta-analysis. A crossroad model. Diplomarbeit, Universität Mannheim.

*Zaridze, D., Maximovitch, D., Zemlyanaya, G., Aitakov, Z. N., & Boffetta, P. (1998). Exposure to environmental tobacco smoke and risk of lung cancer in non-smoking women from Moscow, Russia. *International Journal of Cancer*, 75, 335-338.

*Zhong, L., Goldberg, M. S., Gao, Y. T., & Jin, F. (1999). A casecontrol study of lung cancer and environmental tobacco smoke among nonsmoking women living in Shanghai, China. *Cancer Causes Control*, 10, 607-616.