



Gesamtstellungnahme des WBQ zur NORAH-Studie - Modul 1 (Belästigung und Lebensqualität)

Schriftleitung: PD Dr. Mark Brink, Prof. Dr. Joachim Vogt

Wirkungen von Verkehrslärm auf die Lärmbelästi- gung und Lebensqualität

Autoren der Projektbearbeitung:

Dirk Schreckenber	ZEUS GmbH, Zentrum für angewandte Psychologie, Um- welt- und Sozialforschung, 58093 Hagen
Frank Faulbaum	SUZ - Sozialwissenschaftliches Umfragezentrum GmbH, 47051 Duisburg
Rainer Guski	Ruhr-Universität Bochum, 44780 Bochum
Lars Ninke	SUZ - Sozialwissenschaftliches Umfragezentrum GmbH, 47051 Duisburg
Christin Peschel	ZEUS GmbH, Zentrum für angewandte Psychologie, Um- welt- und Sozialforschung, 58093 Hagen
Jan Spilski	Universität Kaiserslautern, 67663 Kaiserslautern
Jördis Wothge	Ruhr-Universität Bochum, 44780 Bochum

Inhalt

1	Überblick	3
1.1	Überblick über den Forschungsbericht (Hauptteil)	3
1.2	Übersicht über den Anhang	11
2	Gesamteindruck	12
3	Einzelaspekte	14
3.1	Aufbau und Gestaltung des Berichts, Leserführung	14
3.2	Nachvollziehbarkeit, Plausibilität und Konsistenz der zentralen Annahmen (Hypothesenbildung)	15
3.3	Bewertung der eingesetzten Methoden insgesamt und der gewählten Vorgehensweise zur Gebietsauswahl und Probandengewinnung	15
3.4	Umgang mit dem Problem der geringen Teilnahmebereitschaft	16
3.5	Prüfung und Bewertung der Vollständigkeit und Qualität	18
3.6	Diskussion von Ergebnissen und Schlussfolgerungen und Einbettung ins bestehende Schrifttum	19
4	Empfehlung des WBO	19

Sämtliche in diesem Dokument enthaltenen wissenschaftlichen Texte, Grafiken, Tabellen und sonstigen Inhalte sind urheberrechtlich geschützt. Sie dürfen ohne vorherige schriftliche Genehmigung des Urhebers oder des Herausgebers weder ganz, noch auszugsweise kopiert, verändert, vervielfältigt oder veröffentlicht werden. Eine - auch auszugsweise - Veröffentlichung oder Verwendung dieses Dokumentes ist auch mit Zustimmung von Urheber bzw. Herausgeber grundsätzlich nur unter Angabe der vollständigen Quelle zulässig.

1 Überblick

1.1 Überblick über den Forschungsbericht (Hauptteil)

Der vorliegende Forschungsbericht zum Modul 1 der NORAH-Studie umfasst 649 Seiten. Der Text des Endberichts ist untergliedert in fünf Teile: „Einleitung/ Theorie/ Hintergrund“, einen eigentlichen Theorieteil, der hier „Wirkung von Verkehrslärm“ genannt wird, „Untersuchungsplan und Studiendurchführung“, sowie „Ergebnisse“ und „Diskussion, Fazit“. Die in den einzelnen Teilen erscheinenden Unterkapitel werden mittels durchgängiger Nummerierung von 1 bis 18 gekennzeichnet.

Rund 200 zusätzliche Seiten befinden sich in einem separaten Anhang (siehe Kap. 1.2 in dieser Stellungnahme). Diese Menge an Text widerspiegelt in gewisser Weise auch den eindrucklichen Umfang der NORAH-Studie bzw. des Moduls 1, welches seinerseits wiederum aus sieben z.T. fast unabhängigen Teilstudien (Befragungen zur Lärmbelastigung und Lebensqualität bei verschiedenen Quellen und an verschiedenen Orten, Analyse von Kombinationslärmwirkungen) besteht, die methodisch sauber geplant, ausgewertet und letztlich dokumentiert werden müssen.

Im ersten Teil des Berichts werden in **Kapitel 1** die Forschungsfragen, die im Rahmen von Modul 1 bearbeitet werden sollen, vorgestellt. Als Hauptziel für die Studie wird angegeben, dass bei Anwohnern des Flughafens Frankfurt die Expositions-Wirkungsbeziehungen und ihre Veränderungen bezogen auf

- die erlebte Lärmbelastigung und Aktivitätenstörungen am Tage,
- berichtete, lärmbedingte Störungen des Schlafs in der Nacht,
- die gesundheitsbezogene Lebensqualität, aber auch weitergehende Aspekte wie Wohnzufriedenheit oder allgemeine Schlafqualität

im Zeitverlauf vor und nach Eröffnung der Nordwest-Landebahn (NW-Bahn) am Flughafen Frankfurt (FRA) im Oktober 2011 untersucht werden sollen.

Die als Hauptergebnis zu erwartenden Expositions-Wirkungsbeziehungen sollen getrennt für die Verkehrsträger Flug-, Schienen- und Straßenverkehr sowie in Kombination (Flug- und Straßenverkehr bzw. Flug- und Schienenverkehr) betrachtet werden. Außerdem sollen die Expositions-Wirkungsbeziehungen zum Fluglärm unter Berücksichtigung anderer Verkehrslärmarten am Frankfurter Flughafen mit denen an den drei Vergleichsflughäfen, Berlin-Brandenburg (BER), Köln/Bonn (CGN) und Stuttgart (STR), verglichen werden.

In **Kapitel 2** wird einleitend auf die Wirkung von Verkehrslärm in einer Übersicht eingegangen. Ausführlich erörtert werden nicht-akustische Faktoren, die bei der Entstehung von Lärmbelastungsurteilen eine Rolle spielen. Zur theoretischen Fundierung der Befragungen in Modul 1 wird ein stresstheoretisches Rahmenmodell entwickelt. Umfassend wird auf die Literatur zum sog. „Change effect“ (Änderungseffekt) und zur Wirkung unterschiedlicher sowie kombinierter Verkehrslärmquellen eingegangen.

Im zweiten Teil des Berichts, der mit „Untersuchungsplan und Studiendurchführung“ betitelt ist, werden in **Kapitel 3** die faktoriellen Designs an den einzelnen Standorten, das Vorgehen bei der Stichprobenziehung, die abhängigen und unabhängigen Variablen und deren Messung, sowie Fragestellungen und Hypothesen, und nicht zuletzt, die verwendeten statistischen Methoden beschrieben. Der Teil beginnt mit Erörterungen zur Studienregion Rhein-Main bzw. den Studienregionen im Umfeld der Vergleichsflughäfen. Es wird nachvollziehbar ein Stichprobenplan für die Panelstudie mit drei Erhebungswellen in Frankfurt vorgelegt, der von 7.000 Untersuchungsteilnehmern (in der ersten Panelwelle) ausgeht und eine gewisse Panelmortalität über die zweite und dritte Welle berücksichtigt. Die Fallzahlabeschätzung geschah nach den anerkannten Regeln der Statistik mittels einer Poweranalyse, welche auf den Daten der im Rahmen der sog. RDF-Studie, einer ebenfalls am Flughafen Frankfurt (2005) durchgeführten Belästigungsstudie, basierte.

Durch das Untersuchungsdesign mit Schichtung nach aktueller und künftiger Fluglärmbelastung in Frankfurt bzw. nach Ausmaß der Belastungsänderung (gleichbleibend, Abnahme bzw. Zunahme um mehr als 2 dB im Leq) sollte sichergestellt werden, dass durch die Inbetriebnahme der NW-Bahn neu belastete Gebiete ebenso berücksichtigt werden wie Gebiete, die von Verlagerungen von Flugrouten und sonstigen aktiven Schallschutzmaßnahmen betroffen sind. Die Untersuchungsdesigns an den Vergleichsflughäfen orientieren sich an dem in der Frankfurter Panel-Studie verfolgten Konzept der geschichteten Zufallsziehung gemäß langfristiger Luftverkehrsgeräuschbelastung, jedoch im Falle von Köln/Bonn und Stuttgart ohne das Schichtungsmerkmal „Belastungsveränderung“ (da es sich bei diesen um sog. „Bestandsflughäfen“ handelt). Es schließt sich die Beschreibung der Studiendesigns der Querschnittstudien zum Straßen- und Schienenverkehrslärm und zum Kombinationslärm an. Eine sehr umfangreiche und detaillierte Tabelle (Tab. 3-2) und ein Organigramm (Abb. 1-1) geben einen Überblick über sämtliche Teilstudien und deren faktorielle Untersuchungsdesigns.

Im **Kapitel 4** werden die verwendeten Verkehrslärm-Expositionsmaße und die in den Befragungen erhobenen Variablen beschrieben und tabellarisch ausgewiesen. Sodann werden relevante Wirkungs-Zielgrößen benannt (d.h. abhängige Variablen, etwa Variablen zur Belästigung durch Verkehrslärm, zu berichteten lärmbedingten Schlafstörungen und zur Schlafqualität, zur gesundheitsbezogenen Lebensqualität, zu Aktivitätenstörungen, zu psycho-vegetativen Störungen oder zur Wohnzufriedenheit). Sehr detailliert und theoretisch sehr überzeugend begründet wird auf die umfangreiche Liste der „Co-Determinanten“ bzw. „Co-Prädiktoren“ (gemeint sind Mode-

ratoren, Mediatoren) und potenziellen Confounder eingegangen. Auch die potenziell für Lärm-belästigungsurteile konstituierenden situativen Moderatoren (z.B. Wohnbedingungen, Fensterart und -stellung, Schallschutzmaßnahmen etc.) werden vor dem Hintergrund der existierenden Literatur zu diesen Einflussgrößen im Detail erörtert.

Es folgt im **Kapitel 5** die Beschreibung der Forschungshypothesen, die sich aufteilen in Hypothesen zum a) Zeitvergleich, b) Standortvergleich und c) Quellenvergleich, ferner zur Belästigungswirkung kombinierter Einwirkungen und zu den Zusammenhängen zwischen den Wirkungsbereichen Belästigung und gesundheitsbezogene Lebensqualität. Es werden insgesamt 12 Hypothesen bzw. Fragestellungen aufgestellt, die z.T. gerichtet und z.T., mangels ausreichender Vorerfahrung, ungerichtet formuliert sind. Das Signifikanzniveau wird global bei $p=0.05$ (5 %) angesetzt.

Das Vorgehen zur Adressbeschaffung, Stichprobenziehung, Kontaktaufnahme der für die Erhebungen ausgewählten Personen, Haltung, Weitergabe und Auswertung der Daten beschreiben die Autoren in **Kapitel 6**. Für die dabei zu berücksichtigenden Belange des Datenschutzes wurde ein umfangreiches Datenschutz- und Sicherheitskonzept erarbeitet und mit den für den Datenschutz zuständigen Stellen in den beteiligten Bundesländern abgestimmt.

Die umfänglichen Befragungen in der NORAH-Studie, die unterschiedliche Praxis in den einzelnen Bundesländern bei der Herausgabe von Melderegisterdaten, und die z.T. nur langsam mahlenden Mühlen der Verwaltung erforderten vom Studienteam einiges an Überzeugungsarbeit und administrativem Aufwand zur Beschaffung der für die Stichprobenziehung(en) nötigen Einwohnermeldedaten.

Das Vorgehen bei der eigentlichen Stichprobenziehung („Auswahl der Befragungsteilnehmenden“) wird knapp und für die einzelnen Vergleichsflughäfen getrennt erläutert. In die Stichproben aufgenommen wurden grundsätzlich nur Personen, welche über eine recherchierbare Telefonnummer verfügten (oder eine solche nach Aufforderung in einem vorab verschickten Anschreiben mitteilten), was den Nachteil hat, dass eine so gewonnene Stichprobe nicht die ganze Grundgesamtheit abbilden kann. Tatsächlich konnte höchstens bei 50 % der Personen aus der geschichteten Zufallsziehung eine Telefonnummer oder Email-Adresse zugewiesen werden. Die Autoren ergriffen für die spätere Auswertung diverse Maßnahmen, um die Folgen von Unterdeckung und responsebedingten Verzerrungen abzuschätzen (siehe hierzu auch Kap. 3.4 in dieser Stellungnahme). Es wird im Bericht nicht im Detail erläutert, weshalb man sich auf Telefoninterviews festgelegt hat und nicht allenfalls andere Befragungsmodi für die NORAH-Studie evaluiert hat (z.B. schriftliche oder persönliche Befragungen). Bemerkenswert ist, dass den Autoren die Beschränkungen bei Telefonstichproben wohl bewusst waren: Immer mehr Menschen verzichten auf einen (in öffentlichen Telefonverzeichnissen erscheinenden) Festnetzanschluss. So erhofften sich die Autoren, mittels der Bitte im Informationsschreiben, eine Rufnummer oder alternativ Email-Adresse zur Kontaktaufnahme bekannt zu geben, die Auswahlgrundlage anzureichern. Dies geschah allerdings nicht in der erwarteten Größenordnung.

Im weiteren Verlauf des Methodenteils werden die Befragungszeiträume und die Interviewlängen aufgezeigt, sowie auf das Datenmanagement und die Aufbereitung der Daten und detailliert auch auf die statistischen Analysemethoden eingegangen (**Kapitel 7**). Vorbildlich handhaben die Autoren den Umgang mit fehlenden Werten. Durch die große Anzahl verwendeter Multi-Item Skalen in den Befragungen (etwa zur Lärmempfindlichkeit, um nur ein Beispiel zu nennen) sind insbesondere fehlende Werte in einzelnen Items solcher Skalen zu vermeiden, weil sonst die ganze Skala unbrauchbar würde. Deshalb wurden bei Skalenwerten nach Prüfung der Voraussetzungen fehlende Werte, wo immer dies möglich war, mittels differenzierter Imputationsverfahren ersetzt.

Der wichtigste Schritt der Datenanalyse, die inferenzstatistische Prüfung der eingangs formulierten Hypothesen (bzw. die Erstellung von Expositions-Wirkungsmodellen), wird durchdacht und hinsichtlich der verwendeten Analysemodelle begründet. Zur Analyse werden generalisierte lineare Modelle, Strukturgleichungs- und Wachstumsmodelle herangezogen.

Die Frage der Prüfung der Generalisierbarkeit der Ergebnisse und der Repräsentativität der ganzen Studie wird in zwei umfangreichen Unterkapiteln (Kapitel 7.5 und 7.6) behandelt. Aufschlussreich wird (grafisch) gezeigt, auf welchen Ebenen Datenausfälle durch Unterdeckung oder Non-Response auftreten können. Es werden danach die Maßnahmen bzw. Schritte dokumentiert, die zur Prüfung möglicher responsebedingter Verzerrungen vom Studienteam ergriffen wurden (siehe dazu Kap. 3.4 in dieser Stellungnahme) und auch mit welchen Methoden man die Stichproben vor den eigentlichen Auswertungen gewichten könnte, sollte sich dies als nötig erweisen.

Als Ergebnis einer frühen Intervention des WBQ wurden vom Studienteam ab dem Jahr 2012 routinemäßig Non-Response Befragungen durchgeführt. Seit 2012 wurden bei Verweigerern über den Verweigerungsgrund hinaus (der automatisch beim Erstkontakt erfasst wurde) folgende Variablen erhoben: Belästigung durch Flug-, Schienen- und Straßenverkehrslärm, Geschlecht, Geburtsjahr, und höchster Schulabschluss. Leider haben nur lediglich 6 – 10 % der Non-Responder diese Fragen beantwortet. Es war daher dem Studienteam nicht möglich und auswertetechnisch auch nicht sinnvoll, zum statistischen Ausgleich eines Non-Response-Bias eine sog. Non-Response-Gewichtung vorzunehmen.

Der Ergebnisteil beginnt auf Seite 113 mit **Kapitel 8**, zunächst mit einer Übersicht über die Stichprobengewinnung und Übersichten zur Teilnahmebereitschaft an den einzelnen Untersuchungsstandorten. Mittels Flow-Charts wird der Ablauf der Teilnehmerge Gewinnung beschrieben (Abb. 8-1 bis 8-7). In Tab. 8-1 findet man für alle Teilstudien eine detaillierte Aufstellung der Responsestatistik. In Frankfurt ist die Responserate insgesamt unerwartet tief und deutlich niedriger als an den Vergleichsflughäfen. Die Autoren stellen an diesem Punkt deshalb in Aussicht, umfassende Nonresponse-Analysen durchzuführen um den potentiellen Einfluss responsebedingter Verzerrungen auf die Resultate abzuschätzen. In einem ersten Schritt hierzu werden in Kap. 8.5 die Verweigerungsreaktionen analysiert. Sensitivitätsanalysen zur Abschätzung

der Robustheit der statistischen Wirkungsmodelle folgen dann in Kapitel 8.6. (s.u.). Sodann werden die Stichproben der Untersuchungsregionen Rhein-Main am Flughafen Frankfurt sowie an den Vergleichsflughäfen Berlin-Brandenburg, Köln/Bonn und Stuttgart beschrieben. Mehr oder weniger gleich verfahren wird mit den Querschnittstudien (Straßen- und Schienenverkehrslärm, Kombinationsbelastung). Die Darstellung der Verteilungsstatistiken wird ergänzt um inferenzstatistische Angaben zu etwaigen Verteilungsunterschieden. So hat sich gezeigt, dass die Teilnahme an einer NORAH-Befragung z.T. vom Hauseigentum, vom Geschlecht und vor allem vom Befragungsmodus (telefonisch vs. online) beeinflusst wird. Solche Verteilungsunterschiede entsprechen durchaus der Erwartung und die Autoren berücksichtigten in den späteren Analysen den Befragungsmodus, um dessen potenziellen Einfluss auf die Zielvariablen kontrollieren zu können.

Kapitel 9 fällt sehr umfangreich aus und beschäftigt sich mit den zentralen Fragestellungen zur Wirkung des Fluglärms in Frankfurt (FRA) 2011 bis 2013 sowie an den Vergleichsflughäfen Berlin-Brandenburg (BER), Köln/Bonn (CGN) und Stuttgart (STR). Ein erstes Unterkapitel zeigt deskriptive Statistiken zur Luftverkehrsgeräuschexposition in Frankfurt (sowie deren Veränderung über die drei Erhebungswellen) und an den Vergleichsflughäfen, sowie zur Verteilung der Antworten („überhaupt nicht“, „etwas“, „mittelmäßig“, „stark“, „äußerst“) auf die ICBEN-Belastungsfrage insgesamt und aufgeschlüsselt in 5 dB breite Pegelklassen. Die 24h-Mittelungspegel der Panel-Stichprobe lagen in der ersten Welle zwischen 36 und 61 dB(A), in der zweiten und dritten Welle zwischen 35 und 71 dB(A), mit Mittelwerten um 48 dB(A). Die hohen Expositionsstufen in der zweiten und dritten Erhebungswelle sind allerdings nur mit einzelnen Personen besetzt.

Es folgen Ergebnisse inferenzstatistischer Prüfungen (F-Tests) zur abhängigen Variable „Fluglärm-belastigung“ für Haupteffekte und 2-fache Interaktionen von Erhebungswelle, Luftverkehrsgeräuschpegel-Kategorie (in 5 dB-Schritten) und Änderungen in der Luftverkehrsgeräuschbelastung zu den verschiedenen Erhebungszeiten. In allen Wellen steigt die Fluglärmbelastigung mit zunehmendem 24-Stunden-Mittelungspegel $L_{pAeq,24h}$ an. Die Fluglärmbelastigung fiel 2012 insgesamt höher aus als 2011, es zeigte sich allerdings ein Interaktionseffekt zwischen Erhebungswelle und Änderung: Bei Abnahme der Belastung fällt die Belastigung in 2012 etwas geringer aus als in 2011, bei Zunahme dagegen deutlich höher als in 2011 (die Ergebnisse von detaillierten inferenzstatistischen Modellen mit dem Pegel als kontinuierlichem Prädiktor werden im Endbericht weiter hinten präsentiert).

Es folgen Auswertungen zur tageszeitspezifischen Belastigung. Werktags werden höhere Belastigungswerte für die Zeit von 18 bis 23 Uhr berichtet als zu den übrigen Tageszeiten. Es zeigte sich auch, dass die Belastigung durch Fluglärm an Wochenenden höher ausfällt als werktags. Der Vergleich zwischen FRA mit BER, STR und CGN zeigt, dass die Fluglärmbelastigung je Pegelklasse in Frankfurt am höchsten und in Stuttgart am niedrigsten ist.

Im Ergebnisteil des Endberichts kommen auch die Auswertungen zu berichteten, fluglärmbedingten Schlafstörungen nicht zu kurz. Die Auswertungen werden zum einen kategorial mit 5 dB breiten Klassen sowie später mit dem Pegel als kontinuierlicher Variable in erweiterten Modellen durchgeführt. In allen Erhebungswellen konnten die Autoren einen Anstieg der berichteten fluglärmbedingten Schlafstörungen mit zunehmendem Nachtmittelungspegel ($L_{pAeq,22-06h}$) feststellen.

Bei der gesundheitsbezogenen Lebensqualität, die mit dem SF-8 erhoben wurde, wurde vom Studienteam auswertetechnisch sehr ähnlich verfahren wie bei den Outcomes Belästigung und Schlafstörungen. Als wichtigstes Ergebnis kann festgehalten werden, dass die Bedeutung des Fluglärmpegels für die gesundheitsbezogene Lebensqualität im Ausmaß insgesamt wesentlich geringer ist als es der Pegel bei der Fluglärmbelästigung oder den berichteten fluglärmbedingten Schlafstörungen ist. Zusammenfassend ließ sich dennoch feststellen, dass die erlebte psychische Lebensqualität in allen drei Erhebungswellen in höheren Pegelklassen leicht geringer war als in den niedrigeren Pegelklassen.

Nachdem der Bericht an dieser Stelle die Domäne der deskriptiven Darstellungen verlässt, beschäftigt er sich zunächst mit den notwendigen Vorarbeiten zu den später berechneten inferenzstatistischen Expositions-Wirkungsmodellen, wobei von den Autoren zuerst detailliert dargelegt wird, welche Variablen als Confounder, Moderatoren, Mediatoren, oder als Effektmodifikatoren später in die statistischen Modelle zwecks Adjustierung aufgenommen werden können bzw. sollen. Es wird hiernach detailliert beschrieben, mit welchen Methoden Co-Prädiktoren für den Einbezug in die Modelle ausgewählt bzw. verworfen werden. Besonders viel Platz wird den Mediator- und Moderatoranalysen eingeräumt, die nach den neuesten best-practice Prozeduren durchgeführt wurden. Basierend auf diesen Voranalysen werden danach die in die Modelle aufzunehmenden Prädiktoren festgelegt. Das von den Autoren hier gewählte Vorgehen kann als ‚state of the art‘ bezeichnet werden.

In den folgenden Kapiteln (9.7, 9.8, 9.9) schließlich werden die inferenzstatistischen Analysen zu Expositions-Wirkungsmodellen für Belästigung, berichtete fluglärminduzierte Schlafstörungen, und gesundheitsbezogene Lebensqualität beschrieben. Für den Vergleich der Expositions-Wirkungsbeziehung zwischen den Flughäfen bezogen auf die Fluglärmbelästigung in Abhängigkeit des quellenspezifischen 24-Stunden-Mittelungspegels ($L_{pAeq,24h}$) wurden verschiedene Modelle für (A) die Fluglärmbelästigung im Wertebereich von 1 (überhaupt nicht gestört) bis 5 (äußerst gestört oder belästigt) und (B) den Anteil der hoch fluglärmbelästigten Personen (%HA, gemäß IC BEN-Standard) berechnet. Hierbei wurde jeweils ein Basismodell erstellt mit den Prädiktoren Fluglärmbelästigung und Befragungsmodus, sowie ein erweitertes Modell zusätzlich mit den oben beschriebenen bzw. ausgewählten Prädiktoren. In einem Gesamtmodell wurde zusätzlich die Interaktion zwischen Pegel und Flughafen einbezogen. Während im Endbericht die Expositions-Wirkungsbeziehungen grafisch in gewohnter und angemessener Form

in vielen gut aufbereiteten Plots wiedergegeben werden, haben die Autoren die berechneten Modellparameter im Anhang (Anhang 6) platziert.

Die berechneten Modelle bzw. Expositions-Wirkungsbeziehungen zeigen für alle Erhebungswellen in Frankfurt und auch für die Vergleichsflughäfen die erwartete Form. Auffällig ist jedoch – gerade im Vergleich mit der bestehenden Literatur – der generell sehr steile Anstieg der Regressionsgeraden bzw. -kurven sowohl in den linearen (durchschnittliche Belästigung auf der 5-Punkte Skala) als auch den logistischen Modellen (%HA). So werden bei einem durchschnittlichen 24-h-Pegel von 60 dB(A) in Frankfurt um die 80 % hoch Belästigte prognostiziert, mit nur geringfügig tieferen Werten bei den anderen drei Flughäfen. Geradezu symptomatisch bestätigt sich auch der bei vielen ähnlichen Studien beobachtbare Rückgang der Anteile stark Belästigter und stark Schlafgestörter in den höchsten Pegelklassen (vor allem in Frankfurt 2011 sichtbar), ein Phänomen, welches möglicherweise auf Selbstselektionseffekte zurückgeführt werden kann.

Die Belästigungswerte in Frankfurt sind im Durchschnitt über alle drei Befragungswellen höher als an den Vergleichsflughäfen. Im Vergleich mit den anderen Lärmquellen Straße und Schiene wird Fluglärm bei gegebenem 24-h L_{eq} als deutlich belästigender angegeben.

Die Modellierungen für die selbstberichteten fluglärmbedingten Schlafstörungen, die auf dieselbe Weise wie die Belästigung, nämlich mit einer 5-Punkte Skala erhoben wurden, erfolgen analog, mit dem nachvollziehbaren Unterschied, dass die fluglärmbedingten Schlafstörungen auf den Nachtpegel ($L_{pAeq,22-06h}$) bezogen wurden. Auch das Ausmaß der berichteten, fluglärmbedingten Schlafstörungen nimmt, wie erwartet, mit steigender Belastung zu. Die Kurven erreichen jedoch nicht die markanten Steigungen, die bei der Belästigung beobachtet werden können.

Schließlich werden auch Expositions-Wirkungsbeziehungen für den Outcome „gesundheitsbezogene Lebensqualität“ präsentiert. Die in verschiedenen Grafiken dargestellten Mittelwerte zeigen eine leichte Minderung in der psychischen und körperlichen Lebensqualität mit zunehmendem $L_{pAeq,24h}$, was aber insgesamt nicht als dramatisch bezeichnet werden kann. Interessant wäre es an dieser Stelle gewesen, anhand der Altersverteilung in jeder Pegelklasse einen aus der Normstichprobe berechneten „erwarteten“ alterskorrigierten Mittelwert für die jeweilige Pegelklasse zu berechnen und diesen Wert mit den Werten in den einzelnen Pegelklassen zu vergleichen. Damit hätte man letztlich abschätzen können, ob bzw. in welchen Pegel-Klassen Fluglärmbelastung tatsächlich zu einer als relevant zu taxierenden Verschlechterung der Lebensqualität führt.

In **Kapitel 10** schließlich werden vertiefende Längsschnittanalysen zur Wirkung von Fluglärm am Flughafen Frankfurt vor und nach Eröffnung der Nordwest-Landebahn (2011 bis 2013) vorgestellt, die mittels latenten Wachstumskurvenmodellen (LGCM) berechnet wurden. Hierbei wird insbesondere auf den „Change-Effekt“ eingegangen, also auf die Frage, ob nach einer Belastungsänderung „überproportional“ stärkere (im Fall einer Zunahme) oder weniger starke (im Fall einer Abnahme) Belästigungsreaktionen (bzw. fluglärmbedingte Schlafstörungen) zu ver-

zeichnen sind, als in einem beständigen Betrieb. Im Wesentlichen scheint der vermutete „Überschuss“-Effekt eingetreten zu sein, d.h. im Zeitvergleich 2011 – 2013 zeigt sich bei den Befragten der Panelstichprobe, die alle Erhebungswellen mitgemacht haben, im Jahr 2012 ein Anstieg der Belästigung (gegenüber 2011), gefolgt von einem Abfall im Jahr 2013. Das Belästigungsniveau im Jahr 2013 bleibt aber etwas höher als im Jahr 2011. Dieses Muster entspricht exakt dem vorhergesagten und in Abb. 2-4 im Endbericht dargestellten Reaktionsverlauf. Der Anstieg im Jahr 2012 kann als Änderungseffekt interpretiert werden, zumal die Eröffnung der NW-Landebahn eine Reihe von Flugroutenänderungen zur Folge hatte. Der bei der Lärmbelästigung ersichtliche Änderungseffekt zeigt sich bei den berichteten, fluglärmbedingten Schlafstörungen nicht im vergleichbaren Ausmaß. Diese Beobachtung deckt sich auch mit Ergebnissen aus der früheren Literatur über Änderungsstudien. Möglicherweise spielte auch die Einführung des Nachtflugverbots hier eine Rolle.

Zur weiteren Untersuchung der zeitlichen Entwicklung der Fluglärmbelästigung am Frankfurter Flughafen im Zuge der Ausbauplanungen zur NW-Landebahn wurden die Expositions-Wirkungskurven zum Anteil stark fluglärmbelästigter Personen aus der NORAH-Studie mit Ergebnissen aus der RDF-Belästigungsstudie aus dem Jahr 2005 verglichen. Erkennbar ist, dass der Anteil der „stark Belästigten“ (HA) 6 bis 8 Jahre nach der RDF-Belästigungsstudie bei gegebenem Pegel deutlich höher liegt und dass der Unterschied im HA-Anteil zwischen den beiden Studien auch größer ist als die Differenz in den Expositions-Wirkungskurven vor (2011) und nach (2012, 2013) Eröffnung der NW-Landebahn. Dies ist ein weiterer deutlicher Hinweis für den schon seit längerem beobachteten Trend zunehmender Lärmbelästigung bei gegebenem Fluglärmpegel. Für die insgesamt höhere Fluglärmbelästigung der NORAH-Studie im Vergleich zur einige Jahre zurückliegenden RDF-Studie werden von den Autoren zwei unterschiedliche Erklärungen in Betracht gezogen: Es könnte ein eigentlicher „Change-Effekt“ vorliegen, oder das Auseinanderlaufen von Mittelungspegel und Belästigung könnte ein weiteres Beispiel für das sog. „Airport noise paradox“ sein, wonach die Validität des Mittelungspegels als Prädiktor von Belästigungen über die Jahre nachgelassen hat, weil dieser die gestiegene Flugverkehrsmenge nicht angemessen reflektiert.

Die **Kapitel 11 und 12** widmen sich den Untersuchungen zum Straßen- und Schienenverkehrslärm. Es werden jeweils die Expositionen, Zielvariablen, und Expositions-Wirkungsbeziehungen zu Belästigung, berichteten Schlafstörungen und gesundheitsbezogenen Lebensqualität beschrieben. Besonders interessant ist indes **Kapitel 13**, welches die bisherigen Einzelergebnisse der verschiedenen Quellen zusammenführt und einen direkten Vergleich der Lärmwirkungen ermöglicht. Sehr aufschlussreich ist in diesem Zusammenhang Abb. 13-1: Überaus deutlich tritt die beim Fluglärm gegenüber den beiden anderen Lärmquellen wesentlich ausgeprägtere Expositions-Belästigungsbeziehung zutage.

Kapitel 14 beschreibt die bisher im Bericht nicht näher betrachteten Variablen zu Störungen der Kommunikation, zur Ruhe, zu psycho-vegetativen Störungen und zum Lärmbewältigungsver-

mögen, zur selbstberichteten Schlafqualität und Wohnzufriedenheit sowie zu den Veränderungen in der Wahrnehmung (Einstellung zum Flugbetrieb, Vertrauen in die Verantwortlichen und prozedurale Fairness) anhand ihrer Mittelwerte in Bezug auf die entsprechenden Lärmpegelklassen.

In **Kapitel 15** werden die Expositions-Wirkungsanalysen zur kombinierten Verkehrsgeräuschexposition vorgestellt, getrennt für die Kombination Luft-Straße bzw. Luft-Schiene. Insgesamt zeigt sich, dass die Urteile der Befragten zur Gesamtgeräuschbelastung in beiden Substichproben signifikant sowohl mit dem energetisch addierten Gesamtgeräuschpegel ($L_{pAeq,24h}$) als auch dem entsprechenden Einzelpegel für Luftverkehrsgeräusche allein korrelieren, während dies für die jeweils zweite Quelle (hier: Straßen- oder Schienenverkehrslärm) nicht immer gilt. Die Wirkung von Flug- kombiniert mit Straßen- bzw. Schienenverkehrslärm auf die Gesamtbelastung schließlich wird durch energetische Summation der beiden Quellenpegel unterschätzt und besser vorhergesagt durch eine Gewichtung der beiden Pegel entsprechend der Lästigkeit der beteiligten Quellenarten.

In **Kapitel 16** werden die wegen der geringen Response im Frankfurter Panel notwendig gewordenen Abschätzungen potenzieller Ergebnisverzerrung mittels systematischer Sensitivitätsanalysen beschrieben (siehe hierzu detailliert auch Kap. 3.4 in dieser Stellungnahme).

Im letzten Teil, dem Diskussionsteil (IV, beginnend mit **Kapitel 17**) werden die Ergebnisse im Hinblick auf die eingangs gestellten Untersuchungsfragen bzw. Hypothesen diskutiert, rekapituliert und in den bisherigen Erkenntnisstand eingeordnet. Eine Auseinandersetzung mit Stärken und Schwächen der Studie als Ganzes sowie eine Diskussion über deren Limitationen findet in den Kapiteln 17.8 und 17.9 statt.

Der Bericht endet mit **Kapitel 18** im welchem die Autoren nochmals konzis auf die wichtigsten Ergebnisse sowie auf mögliche künftige Entwicklungen eingehen.

1.2 Übersicht über den Anhang

Der 217-seitige Anhang ist untergliedert in eine Darstellung zu „Voranalysen und ergänzenden Sensitivitätsanalysen“, sowie einen Tabellenteil. Die Analyse fehlender Werte und deren Imputation werden im Anhang 1 erörtert. Anhang 2 beschäftigt sich mit der Prüfung der psychometrischen Güte der in den Fragebögen verwendeten Konstrukte (bzw. psychometrischen Skalen). In Anhang 3 werden zusätzliche Sensitivitätsanalysen beschrieben, die nicht im Hauptteil enthalten sind, namentlich zum Vergleich des Effekts verschiedener akustischer Kenngrößen auf die Lärmbelastung, auf berichtete Schlafstörungen und auf die gesundheitliche Lebensqualität. Während die Anhänge 4 und 5 rein deskriptive Verteilungsstatistiken wiedergeben, beinhalten die Anhänge 6 bis 10 in Tabellenform die Koeffizienten der statistischen Regressionsmodel-

le. Anhang 11 schließlich zeigt Tabellen zur Fluglärmbelästigung in Frankfurt zu verschiedenen Tageszeiten.

2 Gesamteindruck

Modul 1 der NORAH-Studie ist eine sozio-akustische Studie zu Fluglärmwirkungen im Raum Frankfurt und an mehreren Vergleichsflughäfen, in welcher eine Reihe von Lärmwirkungen, insbesondere Belästigung, fluglärminduzierte Schlafstörungen und die gesundheitsbezogene Lebensqualität mit Hilfe von telefonischen Interviews und Online-Befragungen untersucht wurden. In der Studie wurde ein sehr umfangreicher Fragebogen eingesetzt, in welchem Angaben zur Person, zu den vermuteten Wirkungen des Lärms, aber auch zu effektmodifizierenden Einflüssen, personalen Faktoren, Bewältigungsstrategien, Gesundheitsverhalten etc. erfasst wurden. Dies hat naturgemäß zur Folge, dass die Menge an beantwortbaren Forschungsfragen und untersuchbaren (psychologischen) Wirkungs-Modellen überaus groß ist und das schlägt sich auch im Umfang des Endberichts nieder.

Die NORAH-Studie (Modul 1) unterscheidet sich von bisherigen (kleineren) Lärmbelastigungsstudien in mehreren Belangen durch wichtige Details. Hervorzuheben sind insbesondere:

- Die Befragung (Frankfurt) in mehreren Erhebungswellen innerhalb eines Änderungskontextes (Interventionsstudie), was erlaubt, Änderungseffekte direkt bzw. in einem quasi-experimentellen Setting (vorher/nachher) zu untersuchen bzw. zu quantifizieren.
- Den gleichzeitigen Einbezug von anderen Flughäfen (Multi-Center-Studie), was die Möglichkeit eröffnet, etwas über die Generalisierbarkeit von Ergebnissen aus ortsspezifischen Lärmwirkungsstudien zu erfahren.

Der Untersuchungsplan berücksichtigt die Kontrolle aller nach dem Stand der Erkenntnis möglichen Confounder, und die Datenanalyse wurde mittels differenzierter statistischer Verfahren durchgeführt. Die Datenerhebungsinstrumente, Testverfahren und Fragebögen wurden auf der Grundlage der theoretischen Vorüberlegungen und Untersuchungsziele sorgfältig ausgewählt und begründet. Die in den Interviews verwendeten Testverfahren sind psychometrisch validiert und in mehreren Sprachversionen verfügbar (z.B. SF-8), was Vorteile hat bei einem allenfalls später angestrebten Vergleich der NORAH-Studie mit anderen Studien bzw. bei der allfälligen Aufnahme der Studie in eine internationale Meta-Analyse.

Als auffällig können die ausgeprägten Belastungs-Belästigungsbeziehungen an allen Flughäfen bezeichnet werden. In Stuttgart beträgt die Korrelation 0,59 was über 30 % erklärte Varianz der Belästigung durch den Pegel bedeutet. Korrelationen in dieser Größenordnung weisen auf mittlere bis starke Zusammenhänge mit der Exposition hin, und weil viele andere Confounder kontrolliert wurden, kann man von mittleren bis starken Effekten des Lärms sprechen. Solch hohe Korrelationen kann man zwar nicht durch besonders sorgfältiges methodisches Arbeiten gezielt

hervorrufen (denn die Enge der bivariaten Beziehung zwischen Belastung und Belästigung ist letztlich eine Eigenschaft der Grundgesamtheit, die man zwar messen, aber nicht beeinflussen kann), sie sind aber dennoch hochgradig erwünscht wenn es darum geht, Lärmschutzmaßnahmen zu begründen, und dies gelingt umso besser, je enger der Zusammenhang zwischen Belastung und Wirkung ist.

Es zeigte sich in den Analysen auch, dass schon vergleichsweise geringe Pegel mit (hohen) Belästigungswerten assoziiert sind, die sich kaum in bisher publizierten Belästigungsstudien finden lassen. Allerdings muss einschränkend dazu angemerkt werden, dass hierfür möglicherweise selektive Response zumindest teilweise verantwortlich sein könnte, insofern als anteilmäßig mehr lärmbelästigte Personen an der Studie teilgenommen haben könnten, als in der Grundgesamtheit vertreten sind (siehe hierzu auch die Ausführungen weiter unten).

Die Rolle nicht-akustischer Faktoren (z.B. Erwartung an den zukünftigen Flugbetrieb, Transparenz in der Kommunikation der Akteure etc.) wurde in Modul 1 in bisher einzigartiger Weise untersucht und aufgeklärt. Hieraus ergeben sich Hinweise auf Lärmbelastungsprävention, die über aktiven und passiven Schallschutz hinausgehen.

Die praktische Anwendbarkeit der Ergebnisse aus Modul 1 ergibt sich aus Sicht des WBQ durch die Verfügbarkeit aufdatierter Expositions-Wirkungsbeziehungen für mehrere deutsche Flughäfen sowie andere Lärmquellen. Die Expositions-Wirkungsbeziehungen für Fluglärm sind nicht nur ausgeprägter als früher, sie unterscheiden sich auch je nach Flughafen, Wirkungszeitraum und Wirkungsart. Zum einen werden damit unterschiedliche Schutzbedarfe deutlich, zum anderen gelten die Expositions-Wirkungsbeziehungen jeweils nur spezifisch und sind nicht ohne weiteres übertragbar (damit ist auch eine grundsätzliche Frage der Lärmwirkungsforschung, nämlich die der Generalisierbarkeit/Extrapolierbarkeit von Modellparametern aus ortsspezifischen Lärmwirkungsstudien beantwortet).

Durch die Verfügbarkeit einer großen Zahl von standardisierten Belästigungsdaten (erhoben mittels ICBEN 5-Punkte Skala) aus verschiedenen deutschen Flughafenregionen, die in jeweils unterschiedlichem Änderungskontext stehen (Bestandsflughafen vs. Flughafen im Änderungskontext) und unterschiedliche temporale Verkehrsstrukturen aufweisen (Nachtflugverbot in Frankfurt, Nachtflüge in Köln/Bonn), also eine hohe Variabilität im Charakter der Lärmbelastung aufweisen, werden die Studienresultate auch im Rahmen von späteren Zusatz-Analysen sehr nützlich sein, z.B. wenn die Frage zu einem Trade-off zwischen Mittelungspegel und Anzahl von Ereignissen untersucht werden soll, oder wenn sie zu einem späteren Zeitpunkt im Rahmen von Meta-Analysen mit anderen Daten gepoolt werden.

Das Problem der insgesamt eher geringen Teilnahmebereitschaft, vor allem bei den Befragungen in Frankfurt, wird vom Autorenteam klar erkannt und als solches unmissverständlich im Bericht identifiziert. Bei hoher Teilnahmebereitschaft besteht ein klarer Zusammenhang zwischen der Höhe der Response und der hier auf jeden Fall angestrebten externen Validität (Repräsentativität für die zugrundeliegende Bevölkerung). Eine 100%-Beteiligung der Stichprobe

würde diese unverzerrt abbilden, ist allerdings unmöglich zu realisieren (vergleichbare Studien gehen schätzungsweise von etwa 20 – 30 % teilnahmebereiten kontaktierten Personen aus). Die ergriffenen Gegenmaßnahmen, d.h. Prüfung potentieller systematischer Selektionseffekte und ggf. darauffolgende Korrektur/Gewichtung der Datensätze werden von den Autoren in mehreren aufeinanderfolgenden Kapiteln abgehandelt, was den Nachvollzug der Problemstellung und der eingeleiteten Lösungsansätze erleichtert. Die entsprechenden Limitationen der Generalisierbarkeit werden in der Diskussion angesprochen und auf potenzielle Verzerrungen aufgrund von Unterdeckung und Verweigerung (also Nichtteilnahme nach Aufforderung zur Teilnahme) wird im Detail eingegangen.

Der WBQ begrüßt die großen Anstrengungen der Autoren zur Kontrolle einer potenziellen Verzerrung der geschätzten Wirkungs-Prävalenzen (z.B. Belästigung) durch Non-Response ausdrücklich und ist sich der sehr komplexen Aufgabe bewusst, die sich hier für das Studienteam gestellt hat.

3 Einzelaspekte

3.1 Aufbau und Gestaltung des Berichts, Leserführung

Die Untersuchungen, die insgesamt im Modul 1 zusammengefasst sind, sind mannigfaltig und übertreffen um ein vielfaches die Zahl an Einzelfragestellungen und damit zusammenhängenden abhängigen Variablen wie sie etwa in den anderen NORAH-Modulen untersucht werden. Dies in eine vom Leser bewältig- und für diesen überhaupt überschaubare Form zu bringen, ist sicher nicht einfach. Die Ergebnisse aus Modul 1 werden in einem sehr umfangreichen Bericht und einem Anhang dargestellt, mit der Folge, dass man sich als Leser leicht in den beiden Dokumenten verlieren kann. Zur verbesserten Leserführung und -orientierung wurde deshalb an den Anfang des Berichts eine detaillierte Zusammenfassung gestellt und eine Übersicht über die einzelnen Studienteile bzw. Sub-Stichproben in einer Art „Organigramm“ (Abb. 1-1) präsentiert. Die umfangreichen Unterkapitel des Ergebnisteils werden jeweils durch ein kurzes Übersichtskapitel eingeleitet. Der Ergebnisteil hätte allerdings im Hinblick auf eine noch bessere Lesefreundlichkeit davon profitieren können, wenn jeweils nach Ende eines Kapitels durchwegs Zwischenfazit bzw. kurze Zusammenfassungen eingeschoben worden wären, welche das zuvor Erarbeitete rekapitulieren. Bei einzelnen Kapiteln haben die Autoren dies getan, bei anderen wiederum nicht.

3.2 Nachvollziehbarkeit, Plausibilität und Konsistenz der zentralen Annahmen (Hypothesenbildung)

Die Studie ist klar und logisch konsistent aufgebaut. Die zentralen Annahmen sind theoretisch gut begründet und sinnvoll hergeleitet worden. Die Fragestellungen der Studie sind vor dem Hintergrund der gegenwärtigen Befundlage zu Fluglärmwirkungen und dem Stand der Theorien und Modelle zur Belästigungs- und Störwirkung von Fluglärm klar nachvollziehbar.

3.3 Bewertung der eingesetzten Methoden insgesamt und der gewählten Vorgehensweise zur Gebietsauswahl und Probandengewinnung

Die in Modul 1 eingesetzten Datenerhebungsinstrumente und Interviewfragebögen wurden entsprechend der einleitend dargestellten theoretischen Grundlagen, der Fragestellungen bzw. Hypothesen ausgewählt. Die Datenanalyse wurde mittels statistischer Verfahren durchgeführt, die dem aktuellen Stand der Wissenschaft entsprechen. Bei der Datenauswertung wurden in vorbildlicher Weise Maßnahmen zur Datenergänzung bei fehlenden Daten durch Imputation getroffen und die entsprechenden Verfahren ausführlich dargestellt (Anhang 1).

Die Gebietsauswahl und die Stichprobengewinnung erfolgten unter den Gesichtspunkten der Abbildung unterschiedlicher Fluglärmbelastungen bzw. der prognostizierten Belastungsänderung, sowie unter Beachtung der anzustrebenden statistischen Power. Bei den Anwerbungen der Studienteilnehmer und der Durchführung der Untersuchung wurden die gängigen ethischen Standards berücksichtigt, bei der Datenerhebung und Datenauswertung wurde den Datenschutzgesetzen Rechnung getragen. Die Auswahl der Gebiete in der Rhein-Main Region, in welchen Befragungen des Frankfurter Panels durchgeführt wurden, ergab sich direkt aus der Auftragsbeschreibung. Als Vergleichsflughäfen wählte das Konsortium gezielt einen weiteren Flughafen mit Änderungskontext, aber ohne Nachtflüge (Berlin) sowie einen „Bestandsflughafen“ mit Nachtflugverkehr (Köln/Bonn) und ohne (Stuttgart). Diese Auswahl eröffnet interessante Vergleiche zwischen verschiedenen Flughafen-Typen. Leider machte die (nicht vorhersehbare) Nicht-Eröffnung des neuen Berliner Flughafens dem NORAH-Team unverdienterweise einen Strich durch die Rechnung. Der für Berlin angestrebte Vorher-Nachher-Vergleich fiel damit ins Wasser.

Eine für die Studie zentrale Entscheidung des Konsortiums war es, als Haupterhebungsmodus auf Telefoninterviews zu setzen und nicht z.B. schriftliche Befragungen, basierend auf Registerdaten (und -Adressen) anzustreben. Durch die Tatsache, dass nicht alle Personen in der Grundgesamtheit über einen recherchierbaren Telefonanschluss verfügten, entstand in den Stichproben eine nur noch schwer korrigierbare Unterdeckung, damit einhergehend eine eingeschränkte Repräsentativität des Frankfurter Panels. Dieses Problem wurde noch weiter verstärkt durch die von den Autoren zu Anfang der Studie zu optimistisch eingeschätzte Teilnah-

mebereitschaft, die in Frankfurt schon in der ersten Erhebungswelle unter 10 % lag. Das Frankfurter Panel war darüber hinaus auch von einer ausgeprägten Panelmortalität („Panelausfälle“) betroffen, die dritte Erhebungswelle umfasste dann nur noch 3.508 Personen von ursprünglich 9.244. Allerdings lag auch dieser Wert noch über der Mindeststichprobengröße (N=2.410) für die Variable mit der geringsten erwarteten Effektstärke (hier gesundheitsbezogene Lebensqualität), die die Fallzahlabeschätzung ergeben hatte.

Es ist in den Sozialwissenschaften unstrittig, dass die Teilnahmebereitschaft bei Bevölkerungsbefragungen in den letzten Dekaden abgenommen hat und dass man als Forscher dagegen relativ machtlos ist. Dass sie indes gar so gering sein würde, wie in Frankfurt erlebt, haben auch interne und externe Qualitätssicherer nicht voraussehen können. Fest steht aber, dass die Auswahlgrundlage bei einem brieflichen Anschreiben (mit z.B. einem beigelegten Papierfragebogen) an aus Melderegistern gezogenen Adressen im Vergleich zum Telefonbuch annähernd die Grundgesamtheit abgedeckt hätte und damit zumindest das Problem der Unterdeckung nicht entstanden wäre. Allerdings gab es für das Konsortium auch gute Gründe, die Befragungen telefonisch und nicht postalisch durchzuführen, denn bei selbstausgefüllten Fragebögen sind Ausfälle durch Item-Nonresponse (Nicht-beantworten gewisser Fragen bzw. Items) erfahrungsgemäß viel häufiger als bei Telefon- oder Online-Interviews, wodurch die Möglichkeit der direkten Prüfung der Antworten viel gezielter – z.B. das vollständige Ausfüllen einer Skala – sichergestellt werden kann.

3.4 Umgang mit dem Problem der geringen Teilnahmebereitschaft

Eher unerwartet wurde das Studienteam des Moduls 1 schon in der ersten Erhebungswelle in Frankfurt mit dem Problem der insgesamt ziemlich geringen Teilnahmebereitschaft konfrontiert. Da die Verallgemeinerungsfähigkeit der Befragungsergebnisse durch die insgesamt geringe Response in den drei Erhebungswellen in Frankfurt (und eingeschränkt auch an den Vergleichsflughäfen) in Frage gestellt war, prüfte das Konsortium sämtliche zur Verfügung stehenden Möglichkeiten/ Maßnahmen um einer potenziell verzerrten Abbildung der tatsächlichen Fluglärmeffekte durch selektive Teilnahme wirkungsvoll begegnen zu können. Während die geringe Teilnahmebereitschaft im Modul 1 nicht dem Konsortium zur Last gelegt werden kann (und so auch vom WBQ nicht erwartet wurde), war es doch die Aufgabe des Konsortiums, das potenzielle Ausmaß der Verzerrung durch selektive Response abzuschätzen und ggf. Gegenmaßnahmen zu ergreifen. Die für das Konsortium relevante Frage war, ob, in welcher Richtung und in welcher Größe die eingeschränkte Repräsentativität des Frankfurter Panels die Schätzung der interessierenden Assoziation zwischen Lärmpegelklassen und Outcomes (Lärmwirkungen) verzerrt. Umfangreicher als in der bisherigen Lärmwirkungsforschung vorfindbar wurden daher Sensitivitätsanalysen durchgeführt, um die Effekte möglicher Selektivität auf die Expositions-Wirkungsbeziehungen zu überprüfen. In Modul 1 betrafen diese

1. die Verteilung von Personenmerkmalen über Geräuschpegelklassen und deren Effekte auf Belästigung, berichtete Schlafstörungen und die gesundheitsbezogene Lebensqualität
2. die Effekte der Gewichtung der Panelstichprobe Rhein-Main nach soziodemografischen Merkmalen der Population (anhand externer Datenquellen)
3. die Non-Responder-Analyse zu den Verweigerungsreaktionen
4. die Vergleiche des Effekts verschiedener akustischer Kenngrößen des Fluglärms auf die Lärmbelästigung, berichtete Schlafstörungen und gesundheitsbezogene Lebensqualität
5. die Vergleiche der Expositions-Wirkungsbeziehungen der Teilnehmenden der einzelnen Erhebungswellen mit der vollen Teilnehmerzahl pro Erhebungswelle in der Panelstudie Rhein-Main
6. die Vergleiche der Expositions-Wirkungsbeziehungen zwischen Basis- und erweiterten Modellen (letztere mit Berücksichtigung nicht-akustischer Faktoren)
7. die Analysen zur Interaktion mit dem Befragungsmodus (telefonisch versus online).

Bei den bevölkerungsbezogenen Daten aus externen Quellen (Punkt 2 in obiger Aufzählung) prüfte das Studienteam zunächst die Qualität der verfügbaren externen Daten und kam zum Schluss, dass nur die soziodemografischen Daten des Bundesstatistikamts von ausreichender Qualität waren, um eine statistische Gewichtung (nach Alter, Geschlecht und Familienstand) der Antwortdatensätze vornehmen zu können. Als Gewichtungsverfahren wurde eine Poststratifikationsgewichtung pro Geräuschpegelklasse durchgeführt. Dabei wurde jede Häufigkeit in der Zelle einer multivariaten Tabelle der Stichprobe (z.B. Geschlecht x Alter x Familienstand) mit den entsprechenden Häufigkeiten in der Population gewichtet. Die möglichen Verzerrungen durch Dropouts im Frankfurter Panel wurden nach Verteilungsprüfung durch eine Propensitygewichtung ausgeglichen. Auf diese Weise wurde die Stichprobe von Welle 2 an die gewichtete Stichprobe von Welle 1 angepasst, die Stichprobe der Welle 3 an die gewichtete Stichprobe von Welle 2. Im Ergebnis zeigte sich, dass die vorgenommene Anpassungsgewichtung der Stichprobe an (a) die ausgewählten Populationswerte und (b) an Merkmalsveränderungen in der Stichprobe aufgrund von Panelausfällen nicht zu einer Veränderung der Expositions-Wirkungsbeziehungen über die Grenzen der Konfidenzintervalle hinaus führte. Die Autoren verzichteten deshalb auf die Gewichtung der Datensätze in den finalen statistischen Modellen.

Obige vom Konsortium in ihrem Auswertekonzept vom 15. Dez. 2014 beschriebenen Prozeduren zur Prüfung allfälliger Non-Response-Verzerrungen im Modul 1 wurden vom WBQ begrüßt und als angemessen befunden. Wenngleich die ergriffenen Möglichkeiten, Responseverzerrungen zu korrigieren, als ausgeschöpft betrachtet werden können – und hierbei dem Konsortium ein wissenschaftlich vorbildliches Vorgehen attestiert werden kann – darf in der Interpretation der modellierten Assoziationen auch mit oder ohne Gewichtung der Fälle nicht per se davon ausgegangen werden, man hätte ein gänzlich unverzerrtes Ergebnis vor sich. Eine Gewichtung von Fällen gemäß der Merkmalsverteilung in der Population kann selbstredend nur für

bekannte Merkmalsverteilungen in der Population erfolgen (z.B. Alter, Geschlecht), während z.B. die Verteilung des Merkmals „Fluglärmbelästigung“ nur in der Stichprobe, nicht aber in der Population bekannt ist. Da wiederum weniger als 10 % der kontaktierten Non-Responder für ein kurzes Non-Response-Interview bereit waren, kam die Möglichkeit einer Non-Response-Gewichtung nicht in Betracht. Ebenfalls nicht durchgeführt wurde eine systematische Vergleichsanalyse zwischen Respondern und Non-Respondern hinsichtlich der relevanten Zielvariablen Belästigung und Schlafstörungen. Ein solcher Vergleich hätte allfällig verbleibende Zweifel ausräumen können, dass sich Responder von Non-Respondern in z.B. ihrer Fluglärmbelästigung nicht systematisch voneinander unterscheiden. Das Konsortium erkannte jedoch klar, dass auch ein solcher Vergleich angesichts der geringen Bereitschaft der Non-Responder für ein Non-Response-Interview 2 – 3 Minuten zu opfern, einer brauchbaren Datenbasis entbehrte – und verzichtete deshalb nachvollziehbar darauf.

Insgesamt können also gewisse residuale Selektionseffekte in den den statistischen Modellen zugrundeliegenden Stichproben auch nach den beschriebenen umfangreichen Prüfungen nicht gänzlich ausgeschlossen werden. Für Modul 1 konnte mit den umfangreichen Sensitivitätsanalysen aber gezeigt werden, dass die dargestellten Expositions-Wirkungsbeziehungen robust sind gegenüber potentiellen Stichprobeneffekten durch die o.g. soziodemografischen Faktoren.

Der WBQ anerkennt ausdrücklich die Tatsache, dass viele ähnlich angelegte sozioakustische Studien wiederum auch von sehr ähnlichen Problemen (insbesondere geringen Responseraten) betroffen sind, geringe Response also keineswegs ein besonderes Merkmal der NORAH-Studie bzw. des Moduls 1 ist (von einem eher geringen Rücklauf war etwa auch die vertiefende Befragung in der Fallkontrollstudie von NORAH-Modul 2.1 betroffen).

3.5 Prüfung und Bewertung der Vollständigkeit und Qualität

Aus den Vergabeunterlagen (dort S. 30) geht hervor, dass die Erarbeitung von Kausalität und Expositions-Wirkungsbeziehungen im Vordergrund stehen, wobei explizit auch gefordert wird, dass, wo dies angebracht scheint, vergleichende Untersuchungen an alternativen Flughafenstandorten durchgeführt werden. Darüber hinaus seien die erzielten Ergebnisse vor dem Hintergrund der aktuellen Lärmwirkungsforschung im In- und Ausland vergleichend zu diskutieren.

Die insgesamt sieben verschiedenen Teilstudien im NORAH-Modul 1 sind den im Angebot formulierten Zielsetzungen gerecht geworden. Die zentralen Ergebnisse sind ausführlich dargestellt und zeigen, dass Modul 1 seine Ziele laut Ausschreibung erfüllt hat, nämlich Kausalitäten und Expositions-Wirkungsbeziehungen im Zeitvergleich, Standortvergleich, Quellenvergleich und bei Kombinationswirkungen zu identifizieren sowie Bezüge zum nationalen und internationalen wissenschaftlichen Schrifttum herzustellen. Dem WBQ sind keine ähnlich umfassenden und/ oder mit ähnlich großen Stichprobenzahlen operierenden Studien bekannt, die zudem mit einer vergleichbar ausgefeilten Methodik ausgewertet worden wären. In diesem Zusammen-

hang besonders lobend hervorzuheben ist der sehr gewissenhafte Umgang mit dem (zu Beginn der Studie nicht unbedingt erwarteten) Problem der geringen Response. Die Um- und Weitsicht, mit welcher hier bei der Auswertung vorgegangen wurde, ist somit selbst ein Qualitätsmerkmal der ganzen Studie. Das Konsortium hat nach Auffassung des WBQ alles (post-hoc) Machbare und Mögliche unternommen, um stichprobenbedingte Verzerrungen aufzudecken bzw. zu korrigieren, namentlich vermittels ausgeprägter Sensitivitätsanalysen (siehe vorheriges Kapitel) und durch die Auswertung der verfügbaren Non-Responder-Informationen, die Prüfung responsebedingter Verzerrung durch Vergleich der Merkmalsverteilung in der Stichprobe und zugrundeliegenden Population, und schließlich durch die Prüfung, ob eine Gewichtung mit Populationsdaten wesentlichen Einfluss auf das Resultat gehabt hätte.

3.6 Diskussion von Ergebnissen und Schlussfolgerungen und Einbettung ins bestehende Schrifttum

Die zentralen Ergebnisse aus dem Endbericht von Modul 1 werden in einer Gesamtschau angemessen dargestellt und sehr gut in den bisherigen Kenntnisstand bzw. die internationale Literatur eingeordnet. Besonderes Gewicht wird hierbei dem Change Effekt gegeben sowie auch die Frage detailliert erörtert, wie die neuen (Frankfurter) Daten sich in den seit mehreren Dekaden beobachtbaren Trend zunehmender Fluglärmbelastigung in der Bevölkerung einreihen.

Gute Forschung muss allerdings nicht nur wissenschaftliche Fragestellungen beantworten, sondern auch neue Hypothesen und Fragen generieren. Daher sind über das Ende von NORAH hinaus weitere Analysen und Validierungsstudien mit Daten aus NORAH allgemein und Modul 1 im Speziellen als Grundlage bzw. Ausgangspunkt angezeigt.

4 Empfehlung des WBQ

Der Wissenschaftliche Beirat Qualitätssicherung (WBQ) hat nach ausführlicher Kommentierung der abgegebenen Entwurfsversion des NORAH-Endberichts Band 3 durch seine Mitglieder die revidierte Fassung vom 07.09.2015 (Version 3) auf seiner Sitzung vom 09.09.2015 angenommen und empfiehlt dem UNH, den Bericht in der vorliegenden Form zu veröffentlichen.