

Ist eine „unbelastete Referenz-Region“ bei Lärmuntersuchungen notwendig?

Angesichts des Umstands, dass das Modul „Entwicklung“ im NORAH-Projekt keine „unbelastete Kontrollgruppe“ untersuchte, hat Prof. Dr. Eberhard Greiser dem NORAH-Team am 21.11.2014 folgende Fragen gestellt:

"In der Epidemiologie ist es seit Jahrzehnten Standard, dass bei der Analyse des möglichen Einflusses eines Risikofaktors z.B. auf das Erkrankungsrisiko von Bluthochdruck nicht nur eine Gruppe von Personen untersucht wird, die diesem Risikofaktor ausgesetzt ist, sondern als Referenzgruppe auch solche Personen in die Studie einbezogen werden müssen, die ohne diesen Risikofaktor leben. Übertragen auf das Problem eines möglichen Einflusses von Fluglärm auf das Risiko an Bluthochdruck zu erkranken, würde man die Häufigkeit von Bluthochdruck in einer von Fluglärm unbelasteten Region mit der Häufigkeit von Bluthochdruck in einer durch Fluglärm belasteten Region vergleichen, weil nur auf diese Weise der Netto-Effekt des Fluglärmeeinflusses abgeschätzt werden kann. Bei Bluthochdruck wären konkurrierende Einflussfaktoren z.B. Salzkonsum, Übergewicht, körperliche Inaktivität. Das Prinzip der Vergleichsregion ist für alle Zustände und Erkrankungen zwingend, wenn neben dem interessierenden Risikofaktor (in diesem Fall Fluglärm) auch andere Risikofaktoren für die Erkrankung oder z.B. Lesefähigkeit von Grundschulkindern in Frage kommen. Wenn neben dem besonders interessierenden Risikofaktor auch andere Risikofaktoren wirken können, muss der Verzicht auf eine Referenzregion zu einer gravierenden Unterschätzung z.B. des Einflusses von Fluglärm führen. Meine Fragen zu dieser Problematik an Sie bzw. die verantwortlichen Kollegen des NORAH-Konsortiums:

Wie haben Sie vor, korrekte Risikoerhöhungen infolge Fluglärms ohne Referenz-Population zu berechnen?

Welche Massnahmen sehen Sie vor, um die Unterschätzung möglicher Fluglärm-Risiken berechnen zu können?"

Antworten des NORAH-Teams:

Idee und Wirklichkeit

Prof. Greiser schreibt in seinen einführenden Sätzen, dass es in der Epidemiologie seit Jahrzehnten Standard ist, bei der Analyse des möglichen Einflusses eines externen Risikofaktors auf ein bestimmtes Erkrankungsrisiko nicht nur eine Gruppe von Personen zu untersuchen, die diesem Risikofaktor ausgesetzt ist, sondern auch eine Referenzgruppe zu untersuchen, die dem Risikofaktor nicht ausgesetzt ist. Dieses Ideal geht auf David Hume (um 1750) zurück, der eine Wirkung als Differenz zwischen dem Ergebnis mit einer bestimmten Behandlung und dem Ergebnis ohne diese Behandlung definierte. Dieses Ideal gilt auch heute noch in der experimentellen Forschung, wenn es z.B. um die Frage geht, ob eine bestimmte (sog. „unabhängige“) Variable (z.B. eine bestimmte akute Geräuschbelastung im Labor) einen temporären Einfluss auf das Verhalten (z.B. Leistungsfähigkeit in einem bestimmten Test) von

Untersuchungspersonen im Labor hat. In solchen Fällen können wir problemlos eine Kontrollgruppe benutzen, die während der Untersuchung der experimentellen Belastung nicht ausgesetzt war.

In der nicht-experimentellen Forschung zu chronischen Wirkungen von Umweltbelastungen ist dieses Ideal jedoch schwer realisierbar, denn wir werden z.B. keine Untersuchungspersonen finden, die niemals irgendeiner Geräuschbelastung ausgesetzt waren. Eine untersuchbare Fragestellung zu Lärmwirkungen muss hier eingeschränkt werden auf die Wirkung spezifischer chronischer Geräuschbelastungen (z.B. Fluglärm oder Straßenlärm oder eine bestimmte Kombination von Geräuschquellen), und das Studiendesign muss die chronische Belastung genau spezifizieren. Hier gehen viele Untersuchungsdesigns den von Stuart Mill (um 1850) gezeigten Weg der Suche nach der „gemeinsamen Kovariation“ von Ursache und Wirkung – hier: akustischer Belastung und Belastungswirkung, und dabei zeigt sich, dass die untersuchten Pegelbereiche in der Realität nicht bis „Null“ hinunter reichen: Beispielsweise kommt in der Meta-Analyse von 14 Untersuchungen zum Zusammenhang zwischen Straßenlärm und koronaren Herzerkrankungen (Babisch, 2014) keine einzige Untersuchung vor, deren niedrigster 16-Stunden-Pegel ($L_{Aeq,16h}$) unter 45 dB liegt. In der internationalen HYENA-Studie (Jarup et al. 2008) wurde für Fluglärm eine untere Pegelklasse von <36 dB $L_{Aeq,16h}$ angesetzt, und die Fluglärm-Untersuchung von Huss et al. (2010) nutzte als niedrigste Pegelklasse <46 dB L_{DN} . Sofern in den Veröffentlichungen eine Begründung für die Wahl der niedrigsten Pegelklasse gegeben wird, werden Unsicherheiten in der Berechnung niedriger Flug- und Straßenlärmpegel genannt – dazu unten mehr. Festzuhalten bleibt, dass eine „unbelastete Kontrollgruppe“ in der internationalen Literatur zu chronischen Lärmwirkungen nicht einmal erwähnt wird und dennoch Expositions-Wirkungsbeziehungen, Odds-Ratios und relative Effekte bei Anstieg der Lärmbelastung um zum Beispiel 10 dB angegeben werden.

Bei bestimmten Fragestellungen ist eine „unbelastete Gruppe“ ohnehin verzichtbar – z.B. dann, wenn der Prozentsatz der Personen mit einer bestimmten Ausprägung bei bestimmten Belastungen gefragt ist, etwa der Prozentsatz hoch belastigter Personen bei X dB Fluglärmpegel. Hierbei wird die erlebte Lärmbelastigung mit Hilfe von Befragungen nach einem standardisierten Verfahren ermittelt, und das Ergebnis kann – bei korrektem Umgang mit „konkurrierenden Einflussfaktoren“ – direkt im politischen und juristischen Bereich genutzt werden. Auch dann, wenn wir uns nicht für die „Gesamtwirkung“ einer definierten Lärmbelastung, sondern (nur) für die Auswirkungen einer spezifischen Intervention – z.B. einer definierten Reduktion des Fluglärm – interessieren, könnte uns ein Studiendesign mit einer „geringer exponierten“ Vergleichsgruppe helfen, und eine „unbelastete Gruppe“ wäre verzichtbar.

Weshalb wird meist keine „durch Fluglärm unbelastete Region“ verwendet?

Die Beschreibung des Ausmaßes der Fluglärmbelastung muss zwangsläufig durch akustische Parameter geschehen. In den oben erwähnten Untersuchungen waren es der 16-Stunden-Pegel ($L_{Aeq,16h}$) und der (gewichtete) Tag/Nacht-Pegel (L_{DN}). Für die Auswahl der Untersuchungsgebiete müssen i.d.R. bereits vorhandene regionale Expositionsdaten benutzt werden. Im Fall der NORAH-Studien waren es $L_{Aeq,06-22h}$

und $L_{Aeq, 22-06h}$ für Fluglärm im Bereich des Frankfurter Flughafens, teilweise verknüpft mit der Festlegung einzubeziehender Verwaltungseinheiten.

Gemeinsam mit den Akustikern des Konsortiums, der internen Qualitätssicherung und des (externen) Wissenschaftlichen Beirats Qualitätssicherung haben wir einen Dauerschallpegel von 40 dB als untere Grenze für die Ermittlung konkreter Lärmpegelwerte bzw. für die Definition der Studienregion für einige Teilstudien, darunter für Modul 3, festgelegt, und zwar aus folgenden Gründen:

An den Messstationen des Flughafens ist auch dann, wenn gar keine Flugverkehrsgeräusche auftreten, immer noch ein Hintergrundgeräuschpegel (L_{95}) von 35 bis 40 dB feststellbar (vgl. NORAH-Akustikbericht für das Kindermodul: <http://www.norah-studie.de/publikationen.epl>). Dieser Pegel kann durch Straßenverkehrsgeräusche, Geräusche aus der Nachbarschaft, von Baustellen, Gewerbebetrieben oder von natürlichen Geräuschen (Wind, Regen) herrühren. Es können zwar auch Fluglärm-Dauerschallpegel unterhalb von 40 dB berechnet werden, aber bei Fluglärm-Dauerschallpegeln unterhalb des „niedrigsten Hintergrundgeräuschpegels“ ist eine quellenspezifische Zuordnung der akustischen Belastung zu festgestellten Wirkungen nicht mehr möglich. Die eindeutige Identifizierung einer Lärmquelle ist aber eine wichtige Voraussetzung für die Fluglärmwirkungsforschung: Es soll ja die spezifische Wirkung des Fluglärms untersucht werden, nicht die Wirkung von Geräuschen aus irgendwelchen nicht definierbaren Quellen. Hinzu kommt der Umstand, dass die Reliabilität von Fluglärm-Dauerschallpegeln mit zunehmender Entfernung vom Flughafen abnimmt. Deshalb wurde die Umhüllende von 40 dB Fluglärm-Tages- oder Nachtpegel als untere Grenze für die Eingrenzung des Untersuchungsgebiets (auf Basis von DES 2007) für einige Teilstudien festgelegt. Die Pegel der adressgenauen Nachberechnungen der Untersuchungsteilnehmerinnen und -teilnehmer für das Jahr 2012 haben sich teilweise erhöht, teilweise verringert. Ausführlicher zu diesem Thema im Hinblick auf das Entwicklungsmodul: <http://www.norah-studie.de/s3/faq.epl> (Frage 9).

Wer unter diesen Umständen eine Fluglärm-Referenz-Region in einem Ballungsgebiet durch einen Dauerschallpegel deutlich unter 40 dB allein auf Grund berechenbarer Fluglärmpegel definiert, muss damit rechnen, dass in dieser Region andere Quellen dominant sind, z.B. Straßen- oder Schienenverkehr oder Gewerbebetriebe. Für Straßen- und Schienenverkehrsgeräusche können, ebenso wie für Fluglärm, adressgenau sehr niedrige Pegel berechnet werden; in statistischen Analysen können diese ggf. als Confounder verwendet werden. Aber Gewerbebetriebe und andere Lärmquellen können damit weder erfasst noch kontrolliert werden. Dazu bräuchte man sehr spezifische lokale Daten. Somit ist allein auf Grund standardmäßig vorhandener Daten weiterhin nicht entscheidbar, welche Geräuschquellenart die „unbelastete Region“ dominiert. Am Ende besteht die Gefahr, dass der Zusammenhang zwischen Fluglärmpegel und Effektgröße in einer Gesamtgruppe analysiert wird, an deren unteren Belastungsende eine undefinierte Quelle oder Quellenkombination steht, während alle anderen Pegelstufen in unterschiedlichem Ausmaß durch Fluglärm dominiert werden. Die faktisch unmögliche Angabe einer quellenspezifischen Fluglärmexposition im Pegelbereich um 40dB führt, ebenso wie andere Unsicherheiten der Fluglärm-Pegelabschätzung im Bereich unterhalb von 40dB, in der Tendenz zu einer konservativen Verzerrung der Effektschätzer. Anders ausgedrückt: Der Einbezug konkreter Schallpegel-Werte unterhalb von 40dB in Regressionsanalysen bzw. die Definition einer Referenzkategorie mit einer

oberen Grenze unterhalb von 40dB kann zu einer systematischen Unterschätzung der Fluglärm-bezogenen Risiken führen.

Was sind „korrekte“ Risikoschätzungen?

Die Expositions-Wirkungsbeziehungen, die wir in einer Stichprobe erhalten, hängen von vielen Faktoren ab – z.B. von der Zusammensetzung der Stichprobe, der Art, dem Range und der Häufigkeitsverteilung der abhängigen und der unabhängigen Variablen sowie deren jeweiliger Reliabilität bzw. Berechnungs-Unsicherheit und Validität. Wir können unter diesen Umständen kein absolutes Korrektheits-Kriterium definieren, sondern müssen uns mit (relativen) Hilfsgrößen begnügen. Dabei erscheint der Vergleich der Ergebnisse zwischen unabhängigen Untersuchungen zur selben Fragestellung als eine wichtige Hilfsgröße, allerdings mit Einschränkungen:

Betrachten wir Expositions-Wirkungskurven im Bereich „kognitive Entwicklung“, sind v.a. die drei Untersuchungen der RANCH-Studie in den Niederlanden, Großbritannien und Spanien relevant (zusammenfassend: Clark et al. 2006). Die drei Untersuchungen benutzten zwar unterschiedliche Fluglärm-Pegelbereiche zwischen 30 und 77 dB, kamen aber jeweils zu sehr ähnlichen adjustierten Expositions-Wirkungsbeziehungen in Bezug auf die Leseleistung. Zwar war diese Leistung bei der niedrigsten Fluglärmbelastung keineswegs am höchsten, was daran denken lässt, dass die lineare Beziehung steiler würde, wenn man die niedrigste Pegelstufe unberücksichtigt lässt, aber statistische Vergleiche zwischen den Untersuchungen zeigen, dass sich die Steigungen der drei Untersuchungen nicht signifikant unterscheiden. Die entsprechende NORAH-Untersuchung kam selbst in dem eingeschränkten Pegelbereich von 39-59 dB Dauerschallpegel zu sehr ähnlichen Ergebnissen.

Wann wird ein Risiko unter- oder überschätzt?

Wie bereits erwähnt, kann der Einbezug konkreter Schallpegel-Werte unterhalb von 40 dB in Regressionsanalysen zu einer Unterschätzung der Fluglärm-bezogenen Risiken führen. Dies liegt an der Unsicherheit und geringen Reliabilität von sehr niedrigen Dauerschallpegeln, die zu Fehlklassifikationen und einer Erhöhung der Fehlervarianz führen. Theoretisch könnte der Einbezug konkreter Schallpegel-Werte unterhalb von 40 dB auch zu einer Überschätzung der Fluglärm-bezogenen Risiken führen, wenn beispielsweise ausgehend von einem (tatsächlich verzerrungsbedingten) „Plateauverlauf“ der Expositions-Wirkungs-Beziehung im Bereich unterhalb von 40 dB die „Steilheit“ des Expositions-Wirkungs-Verlaufs durch die Wahl bestimmter (tatsächlich ungeeigneter) – Funktionen (z.B. bestimmter Polynome höheren Grades) überschätzt würde. Allerdings erscheint uns die Gefahr einer Unterschätzung des Risikos durch den Einbezug konkreter Schallpegel-Werte unterhalb von 40 dB größer, und durch den Verzicht auf niedrigere Schallpegel-Werte möchten wir dieser Gefahr begegnen.

Wir möchten darauf hinweisen, dass wir mit modulbezogen unterschiedlichen Maßnahmen versuchen, auch etwaigen Fluglärmwirkungen im Bereich eines Dauerschallpegels unter 40 dB gerecht zu werden: So legen wir in unserer Fallkontrollstudie (als Teilmodul des Moduls 2) den Cutpoint für die

Referenzkategorie zwar wie oben beschrieben ebenfalls auf einen Dauerschallpegel von 40 dB. Um auch etwaigen relevanten "Lärmereignissen" unterhalb eines Dauerschallpegels von 40 dB gerecht zu werden, schließen wir in einem zweiten Ansatz diejenigen Probanden aus der Referenzkategorie aus, die bei einem Dauerschallpegel unter 40 dB einem nächtlichen Maximalpegel von 50 dB oder mehr ausgesetzt sind. Wenn die beschriebene Aufteilung dieser Referenzkategorie eine deutlich bessere Modellgüte ergibt (AIC mind. 5 Punkte niedriger), wird gemäß unserem a priori festgelegten Auswertungskonzept der zweitgenannten Referenzkategorie der Vorzug gegeben.

Literatur:

Babisch, W. (2014). Updated exposure-response relationship between road traffic noise and coronary heart diseases: A meta-analysis. *Noise & Health*, 16(68), 1-9.

Clark, C., Martin, R., van Kempen, E., Alfred, T., Head, J., Davies, H. W., Haines, M.M., Lopez, B.I., Matheson, M. & Stansfeld, S. A. (2006). Exposure-effect relations between aircraft and road traffic noise exposure at school and reading comprehension: The RANCH project. *American Journal of Epidemiology*, 163(1), 27-37.

Huss, A., Spoerri, A., Egger, M., Röösli, M. & for-the-Swiss-National-Cohort-Study-Group. (2010). Aircraft Noise, Air Pollution, and Mortality From Myocardial Infarction. *Epidemiology*, 21(6), 829-836.

Jarup, L., Babisch, W., Houthuijs, D., Pershagen, G., Katsouyanni, K., Cadum, E., Dudley, M.L., Savigny, P., Seiffert, I., Swart, W., Breugelmans, O., Bluhm, G., Selander, J., Haralabidis, A., Dimakopoulou, K, Sourtzi, P., Velonakis, M. & Vigna-Taglianti, F. (2008). Hypertension and Exposure to Noise near Airports - the HYENA study. *Environmental Health Perspectives*, 116, 329-333.

Bochum, am 15.12.2014. Für das NORAH-Team:

Rainer Guski, Andreas Seidler, Dirk Schreckenber, Ulrich Möhler und Maria Klatte.