

Ausgewählte Fragen und Antworten aus wissenschaftlicher Sicht zur MAK-NO₂-Studie in Aachen

Autor: Rudolf A. Jörres, Institut und Poliklinik für Arbeits-, Sozial- und Umweltmedizin, Klinikum der Universität München, LMU

(1) Wurden die Probanden gesundheitlich gefährdet?

Um Studien experimenteller Art beim Menschen zu verstehen, muss man unterscheiden zwischen

- (a) Effekten wie biochemischen und zellbiologischen Reizwirkungen, geringen Funktionsänderungen usw., die eine bloße vorübergehende Reaktion anzeigen und oft nicht einmal vom Probanden wahrgenommen werden. Solche Effekte sind von Interesse, um Hinweise auf Mechanismen zu erhalten, die bei Konzentrationssteigerung akut sowie bei niedriger Konzentration langfristig klinisch relevant werden könnten.
- (b) Effekte mit dem Potential unmittelbarer klinischer Relevanz. Dies betrifft vor allem spezielle indirekte Wirkungen, z. B. eine Steigerung der Reaktion auf eingeatmetes Allergen beim allergischen Asthma nach vorheriger Inhalation von Luftschadstoffen, die u.U. für sich genommen in der gewählten Konzentration keine direkt messbaren Effekte hervorrufen, z.B. Holz et al. 2002 für Ozon. Solche Effekte könnten im Prinzip bei weiterer Verstärkung den Probanden zu einem Arztbesuch nötigen. Durch eine vorsichtige Dosissteigerung vermeidet man jedoch, dass die experimentellen Effekte klinisch intolerabel und unbeherrschbar werden. Es geht wie fast immer darum, eine „Richtung“ zu identifizieren. Effekte dieser Art sind von Interesse, wenn man eine spezifische Gefährdung spezieller Patientengruppen erfassen möchte, ebenso als Leitlinie, um zu gezielterem Nachfassen in epidemiologischen Studien zu motivieren.
- (c) Dann gibt es natürlich Effekte, die den Probanden tatsächlich gefährden, z.B. schwer beherrschbare Verengungen der Bronchien durch spezielle Substanzen bei Asthmatikern oder gravierende Störungen der Herzfunktion. Solche Effekte sind in allen experimentellen Studien auszuschließen, da sie ethisch nicht vertretbar sind, und es bedarf der Erfahrung und Kompetenz der Untersucher, Design und Probanden so zu wählen, dass sie ausgeschlossen sind.

Für gesunde Probanden und NO₂ im verwandten Konzentrationsbereich sind höchstens Effekte der Kategorie (a) zu erwarten. Anderslautende Behauptungen sind gegenstandslos.

(2) Worum ging es in der Untersuchung?

- (a) Um den neu festgesetzten MAK-Wert von 0,5 ppm. MAK-Werte beziehen sich auf Gesunde, nicht auf Vorerkrankte. Ferner sollten MAK-Werte im Idealfall so gewählt werden, dass überhaupt keine Effekte auftreten, selbst wenn solche im Normbereich bleiben und gar nicht vom Probanden wahrgenommen werden. Eine Reaktion irgendwelcher Art gilt als Hinweis darauf, dass langfristig etwas Abträgliches geschehen könnte. Wohlgedenkt „könnte“, da ist eine große Sicherheitsmarge.

- (b) Der Focus liegt somit etwas anders als bei experimentellen Umwelt-Studien. Bei diesen interessiert naturgemäß ebenfalls, ob die exponierten Personen kurzfristig oder langfristig eine Erkrankung entwickeln. Man ist aber sinnvollerweise eher an einzelnen Markern interessiert, die sich direkt in Richtung spezifischer Erkrankungen oder Verschlimmerungen von Erkrankungen (z.B. Asthma oder Herzerkrankungen) interpretieren lassen. Hierfür dienen die Beobachtungen aus epidemiologischen Studien als Hinweise. In der Regel sind es ja gerade Vorerkrankte, die besonders gefährdet sind. Daher hätte man z.B. bei Patienten mit Asthma oder COPD andere oder zusätzliche Messungen durchgeführt, z.B. solche, die auf allergische Reaktionen und bronchiale Überempfindlichkeit oder auf Störungen des Gasaustauschs abzielen (siehe Frage 13).

Wenn man die Sinnhaftigkeit der Studie beurteilen möchte, sollte man diese Überlegungen parat haben.

(3) Hätte man überhaupt Effekte erwarten können?

Ja, zumindest bei 1,5 ppm, und diese Konzentration war als Positivkontrolle gedacht. Siehe z.B. aus vorangegangenen Arbeiten aus einer der beteiligten Arbeitsgruppen Effekte bei 1 ppm NO₂ (Jörres et al. 1995) sowie fragliche Effekte bei 0,5 ppm (Jörres & Magnussen 1990, Jörres & Magnussen 1991). Behauptungen der Art, es sei von vornherein klar gewesen, dass in der Studie keine Effekte auftreten würden, sind gegenstandslos. Siehe dazu auch die detaillierten Ausführungen zur Messmethodik unter Frage 4.

(4) Welche Effekte hätte man erwarten können, warum wurden welche Methoden eingesetzt?

- (a) Für NO₂ sind im untersuchten Konzentrationsbereich bei Gesunden keine messbaren oder relevanten Änderungen der Mechanik der Atemwege zu erwarten, auch wenn solche bei Asthmatikern angedeutet sein können (z.B. Jörres et al. 1995). Diese Effekte misst man in der konventionellen Lungenfunktion mittels sogenannter Spirometrie und Ganzkörperplethysmographie. Sollten Effekte von NO₂ auf die periphere Lunge auftreten, könnten die sich in kleinen Änderungen des Ruhevolumens der Lungen widerspiegeln, wie man sie in der Ganzkörperplethysmographie erfassen kann. Man muss zumindest die spirometrischen Parameter messen, sonst würde das Fehlen moniert. Auch dient die Messung der Lungenfunktion mit dazu, sicherzustellen, dass die Probanden sich im gleichen klinischen Zustand befinden und gesund sind.
- (b) Die Autoren von Nieding & Wagner (1979) beobachteten bei Patienten mit chronischer Bronchitis mittels Erfassung des Gasaustauschs bei 0,5-8 ppm NO₂ Effekte, die sie in Richtung eines „bronchiolären, alveolären und interstitiellen Ödems“ deuteten; unterhalb von 1,5 ppm waren mit den dort verwandten, relativ groben und indirekten Methoden keine Effekte nachweisbar. In Übereinstimmung damit könnte man spekulativ (bei extremer Extrapolation nach unten) aufgrund der Tatsache, dass bei unfallmäßig hohen Konzentrationen ein regelrechtes Lungenödem auftreten kann, Änderungen des Flüssigkeitshaushaltes der Oberfläche der Alveolen (Lungenbläschen) erwarten. Solche Effekte könnten angedeutet auch beim Gesunden auftreten und wären im Übrigen von besonderem Interesse für eine Extrapolation auf die epidemiologische Situation (siehe

Fragen 9 und 11). Allerdings sollten sie sich bei Gesunden nicht in einer messbaren Störung des Gasaustauschs, wohl aber vielleicht in einer Einschränkung der maximal möglichen Sauerstoffaufnahmekapazität bemerkbar machen.

- (c) Dazu könnte man *prima vista* die klinisch gut eingeführte sogenannte Diffusionskapazität für Kohlenmonoxid (CO) messen. Jedoch sind Änderungen der konventionellen CO-Diffusionskapazität bei Gesunden unwahrscheinlich, da diese vom verfügbaren Blutvolumen der Lunge dominiert wird, für das man bei Gesunden keine Änderungen erwarten wird. Hier springt das spezielle, avancierte Verfahren der kombinierten Diffusionskapazität für CO und eingeatmetes Stickstoffmonoxid (NO) ein. Da die Aufnahme von NO praktisch nur von eventuellen „echten“ Transportstörungen, nicht aber vom Blutvolumen abhängt, sollten damit auch sehr kleine Änderungen messbar sein, wie man sie bei Änderungen des Flüssigkeitshaushaltes der Lungenbläschen erwartet. Bei einer Zunahme der Flüssigkeitsmenge auf der Oberfläche sollte der Übertritt des zu Messzwecken eingeatmeten NO in das Blut erschwert sein. Für diese Methode lagen ausgiebige Erfahrungen in einer der Arbeitsgruppen vor (z.B. Dressel et al. 2009), ferner gibt es inzwischen für eine mögliche Interpretation in diese Richtung Hinweise (Karrasch et al. 2017).
- (d) Es ist immer vernünftig, eine Hypothese mit mehreren Messgrößen zu überprüfen. Ebenso wie der Übertritt von eingeatmetem NO in das Blut erschwert sein könnte, könnte auch der (sehr geringfügige) Austritt von NO aus dem Blut in die Ausatemluft erschwert sein. Hierzu kann man das sogenannte alveoläre NO messen. Dies ist allerdings alles andere als einfach, da die Konzentration in der Ausatemluft von dem in den Bronchien erzeugten NO dominiert wird. Man kann jedoch auf das alveoläre NO aus Messungen bei verschiedenen Ausatemgeschwindigkeiten unter Anwendung eines Modells zurückschließen. In einer der Arbeitsgruppen wurde über viele Jahre an der Entwicklung und Verbesserung dieser Technik gearbeitet; dort liegt Expertise wie sonst nirgendwo in Deutschland vor. Man kann annehmen, dass das alveoläre NO nicht nur durch die Höhe der Barriere für den Übertritt aus dem Blut, sondern auch durch seine Reaktion mit reaktiven Sauerstoffverbindungen im alveolären Raum beeinflusst wird. Da NO₂ pro-oxidativ wirkt, war es sinnvoll, einen möglichen senkenden Effekt auf das alveoläre endogen erzeugte NO anzunehmen. Die Messung des bronchial erzeugten NO (FeNO) ist Standard und messtechnisch mit der Messung des alveolären NO gekoppelt.
- (e) In ähnlicher Weise erlaubt es die Technik des Atemkondensats (EBC), biochemisches Material aus der (peripheren) Lunge zu gewinnen, ohne invasiv zu werden. Hierbei wird die Ausatemluft über 15 Minuten durch ein Kühlaggregat geleitet und die darin enthaltene Feuchtigkeit ausgefroren. Diese wird nicht nur durch Wasserdampf gestellt, sondern auch durch einen kleinen Anteil von Aerosolen, die in der Lunge während der Atmung erzeugt werden und biochemische Komponenten transportieren können, die nicht flüchtig sind und verdampfen wie Wasser. Die gemessenen Verbindungen Nitrit und Nitrat waren auf die Exposition gegenüber NO₂ direkt abgestimmt, einmal als potenzielle Reaktionsprodukte, dann als Produkte einer zellulären Aktivierung.
- (f) Die Technik des induzierten Sputums hat sich vielfach, z.B. bei Expositionen gegenüber Ozon (z.B. Holz et al. 1999) sowie in klinischen Studien als hervorragend geeignet erwiesen, zelluläre und biochemische Antworten wenig invasiv zu messen. Mit ihr

gewinnt man zelluläres und biochemisches Material aus den Bronchien durch Abhusten, das durch Einatmung konzentrierter Kochsalzlösung erleichtert oder hervorgerufen wird. Diese Messung ergänzt die auf die Peripherie der Lunge bezogenen vorher genannten Techniken.

- (g) Ähnliches gilt für die Technik der Gewinnung von Proben des Nasensekrets, mit denen man direkt die Antwort in der Schleimhaut der Nase messen kann. Auch dieses Verfahren verlangt Erfahrung und ist technisch nichttrivial. Die beteiligten Arbeitsgruppen verfügten über solche (z.B. .
- (h) Bleiben Fragen zum Protokoll in Hinsicht auf den zeitlichen Ablauf. In einer vorangegangenen Studie aus einer der Arbeitsgruppen (Jörres et al. 1995) wurden bereits 1 Stunde nach einer dreistündigen Exposition gegenüber 1 ppm NO₂ mit wiederholter Fahrradbelastung kleine, aber statistisch signifikante Effekte auf die Konzentrationen von Prostaglandinen im Bronchialsystem beobachtet. Daher konnte man auch in der vorliegenden Studie Effekte erwarten, zumal die Messungen nach Exposition eine längere Zeit in Anspruch nahmen und die Analyse der Nasenproben und des induzierten Sputums zum Schluss erfolgten. Im Prinzip ist es wünschenswert, längere Zeit nach Exposition zu messen, um verzögerte Effekte zu erfassen, im Idealfall nach 6 oder 12 oder 24 Stunden. Das praktische Problem ist, dass die Probanden, falls sie „aushäusig und außer Kontrolle“ sind und ins Labor zurückkommen, zwischendurch unkontrollierten Belastungen ausgesetzt sind, die die Antwort verzerren, vor allem dann, wenn man in den Expositionen nur vergleichsweise kleine Reize gesetzt hat (man denke nur an aushäusige Expositionen im Straßenverkehr). Wenn man die Probanden stationär unterbringt, steigen die organisatorischen und finanziellen Anforderungen erheblich. Das sollte man bedenken.

Somit deutet eine Kritik an dem Panel der Messungen als möglicherweise inadäquaten Prozeduren auf mangelhafte Expertise bzw. Verständnisdefizite hin.

(5) Warum wurde die niedrige Konzentration von 0,1 ppm in der Studie verwandt?

- (a) Die niedrige, umweltrelevante Konzentration war einerseits als zweiter möglicher „Leerwert“ gedacht, da wir nicht mit Effekten rechneten. Wenn man nahe der vermuteten Nachweisgrenze arbeitet, ist eine stabilere, sozusagen doppelt gemessene Grundlinie als Vergleichswert immer von Vorteil. Andererseits kann man Effekte im Sinne nichtlinearer, U-förmiger Verläufe nie ausschließen, d.h. stärkere Änderungen der gemessenen Marker bei niedrigen Konzentrationen. Nichts war da im Vorab klar, und wer wissenschaftliche Erfahrung hat, weiß aus eigenem Erleben, dass man immer wieder Überraschungen erlebt. Erst die statistische Analyse konnte zeigen, ob bei 0,1 ppm Effekte auftraten oder nicht. Die statistische Teststärke (Power) würde in jedem Fall steigen.
- (b) Zur Veranschaulichung möge das folgende klinische, nicht umweltbezogene Beispiel dienen: Ein bestimmter Typ von Atemwegsentszündung (eosinophil) steht mit einer allergischen Reaktion bzw. mit allergischem Asthma in Verbindung, und als Marker eignet sich das ausgeatmete, bronchial erzeugte NO. Wenn die Reaktion stärker wird,

kann jedoch zusätzlich ein anderer Entzündungstyp (neutrophil) auftreten, der mit der Produktion von reaktiven Sauerstoffverbindungen durch diese Zellen verbunden ist. Diese Verbindungen reagieren teilweise mit dem in der bronchialen Schleimhaut produzierten NO und senken dessen Konzentration. Auch können Schleimhautschwellungen auftreten, die den Übertritt des NO in die Ausatemluft erschweren. Das bedeutet, dass man mittels des Markers des ausgeatmeten NO bei niedrigem Reiz u.U. größere Effekte beobachtet als bei höherem Reiz. Was wir messen können, sind aber immer nur Marker, also mehr oder weniger indirekte Indikatoren. Es ist möglich, dass eingeatmetes NO₂ ebenfalls mehrphasige Effekte auslöst.

(6) War die Fallzahl der Probanden nicht zu gering?

Die erforderliche Fallzahl hängt von den zu messenden Zielgrößen und dem Design der Studie ab, die beide das statistische Auflösungsvermögen bestimmen. Experimentelle Expositionsstudien am Menschen weisen in der Regel bestimmte Designs auf, und zwar aus den folgenden Gründen.

- (a) Die Variationsbreite von Messgrößen innerhalb einer Population von Menschen ist in der Regel viel größer als die Effekte, die man verantwortlicherweise durch eine experimentelle Exposition auslösen kann. Daher gibt es Schwierigkeiten, letztere überhaupt zu erkennen. Zugleich sollte aus Gründen des zeitlichen und materiellen Aufwandes die Zahl der Probanden möglichst realistisch, d.h. klein sein. Daher optimiert man das Design so, dass man kleine individuelle Effekte von den großen Unterschieden zwischen den Probanden möglichst gut trennen kann.
- (b) Das optimale Design umfasst (A) den Ansatz, jeden Probanden mit sich selbst zu vergleichen, indem er einer oder mehreren Konzentrationen von NO₂ sowie einer Kontrolle (Reinluft) ausgesetzt wird. Man wertet für jeden Probanden nur noch die individuellen Änderungen gegenüber der Kontrollexposition; dadurch gehen idealerweise die großen Unterschiede der Messgrößen zwischen den Probanden nicht mehr in die statistische Analyse ein. Es gibt aber noch eine weitere Quelle der Variabilität (B), indem die Probanden von Tag zu Tag leicht unterschiedliche Werte zeigen können. Dieses Problem geht man an, indem man vor und nach Exposition misst und die Differenz beider Werte analysiert.
- (c) Auf diese Weise behandelt man auch das mögliche Problem (C), dass die Werte vor und nach Exposition aufgrund einer natürlichen Änderung über den Tag verschieden sein können, ferner (D), dass bestimmte Messungen vor Exposition selbst einen Effekt auf die Messwerte nach Exposition haben könnten. Solche Gedächtniseffekte sind vor allem evident für invasive Techniken: wenn man einmal Teile der Lungen mit Kochsalzlösung spült, kann man nicht erwarten, dass diese Bereiche nicht durch diese Spülung selbst unabhängig von allen Schadstoffen beeinflusst werden. Effekte der Art (C) und (D) erkennt man an der Kontrollexposition.
- (d) Um zeitliche Übertragungseffekte zu vermeiden, führt man die Expositionen in zufälliger Reihenfolge durch, idealerweise ferner so, dass zumindest die Probanden „verblindet“ sind, indem sie nicht wissen, welcher Exposition sie gerade ausgesetzt sind. Übertragungseffekte können zustande kommen durch (E) unerkannte längerfristige Effekte,

vor allem bei kurzer zeitlicher Abfolge der Expositionen, aber auch (F) durch die zunehmende Erfahrung der Probanden, die vor allem bei komplexen, mit arbeitsabhängigen Messverfahren oder bei der subjektiven Wahrnehmung der Situation von Bedeutung sein können.

- (e) Den daraus resultierenden Ansatz nennt man „controlled, randomized, cross-over, pre-post design“. Er reduziert die Zahl der Probanden erheblich. In vielen Studien haben Probandenzahlen von ca. 20 bis 30 als getrennt analysierte Subgruppen oder als Gesamtzahl genügt, um (G) statistisch signifikante Effekte auf Luftschadstoffe zu erkennen oder (H) in Abwesenheit signifikanter Änderungen der Mittelwerte einzelne Probanden mit klarer Reaktion zu identifizieren (siehe aus den eigenen Arbeitsgruppen z.B. Jörres et al. 1995, Jörres et al. 1996, Holz et al. 1999, Jörres et al. 2000, Holz et al. 2002, Schnuch et al. 2010). Es gab im Vorfeld keinen Grund, anzunehmen, dass für die eingesetzten Messgrößen und Konzentrationen von NO₂ die Situation eine andere sein sollte.
- (f) Epidemiologische Studien, vor allem ein „Screening“ auf arbeitsplatz- oder umweltbezogene Erkrankungen usw., finden unter grundsätzlich anderen Bedingungen statt. Weder ist in der Regel ein und derselbe Proband dem zu untersuchenden Luftschadstoff sowie einer Reinluftkontrolle ausgesetzt, noch besteht die Möglichkeit eines Vor-Nach-Vergleichs. Aus diesem Grunde schlagen die großen natürlichen Unterschiede von Messgrößen zwischen den Probanden voll zu Buche. Selbst wenn man in der glücklichen Lage ist, mehrere Messungen bei denselben Probanden vergleichen zu können, hat man es in der Regel mit wenigen Messgrößen und vor allem mit schlecht definierten Expositionen zu tun. Aus allen diesen Gründen werden immer große Fallzahlen benötigt. Abgesehen davon sind auch die Fragestellungen epidemiologischer Studien andere als diejenigen experimenteller Studien (siehe Fragen 9 und 10).
- (g) Man kann sich die Situation an folgendem fiktiven Beispiel klarmachen. Angenommen man wollte untersuchen, ob ein Medikament das Körpergewicht kurzfristig und vorübergehend geringfügig steigert (z.B. auf Basis des Wasserhaushaltes). Man würde das Medikament sowie ein Placebo geben und die Änderungen nach beiden Interventionen messen und miteinander vergleichen. Jeder Proband dient dabei als sein eigener Bezug. Da andererseits das Körpergewicht in der Population eine sehr große Streubreite zeigt, würde man offensichtlich eine riesige Zahl an Probanden benötigen, um einen Unterschied zu erkennen, wenn man der einen Hälfte das Medikament und der anderen Hälfte Placebo verabreichen wollte. Das sollte auch ohne statistische Expertise intuitiv klar sein.
- (h) Es ist nützlich, sich den außerordentlichen Einfluss des Studiendesigns im Anschluss an (g) an einem orientierenden Zahlenbeispiel vor Augen zu führen. Man wähle einen üblichen statistischen Test auf Basis der Normalverteilung. Für diesen ist entscheidend der Quotient aus (erwarteter) Effektgröße und Standardabweichung. Falls der Quotient gleich einem kritischen Wert von 2 ist, ist eine „Signifikanz auf dem 5%-Niveau im beidseitigen Test“ erreicht.

Wenn die Standardabweichung der täglichen Schwankung des individuellen Gewichts plus/minus 1 kg beträgt (was viel ist) und man einen Effekt von 0,5 kg erwartet, beträgt

der Quotient 0,5 (0,5kg/1kg), d.h. der erwartete Effekt liegt bei 50% der Standardabweichung. Entscheidend für den Test ist aber die Standardabweichung des Mittelwertes, die gleich der Standardabweichung einer einzelnen Messung dividiert durch die Wurzel aus der Zahl der Messungen (in diesem Fall Probanden) ist. Da die Standardabweichung im Nenner des oben genannten Quotienten steht, kommt die Wurzel aus der Zahl der Messungen als Faktor in den Zähler (doppelter Bruch). Für eine Abschätzung der Fallzahl suchen wir somit eine Zahl, deren Wurzel multipliziert mit 0,5 den kritischen Wert 2 ergibt. Diese Zahl ist offenbar 16, denn 4 mal 0,5 ergibt 2. Somit reichen in dieser Abschätzung 16 Probanden bei einem individuellen pre-post-Design.

Gibt man aber das Medikament zwei Gruppen der Population, sieht die Sache ganz anders aus (hierbei ist das Gewicht der Probanden vorher nicht bekannt, analog biochemischen Messgrößen, und kann also auch nicht zum „Matchen“ der Gruppen vorab verwandt werden). Nehmen wir an, in der Population, d.h. über alle Personen gesehen, betrage die Standardabweichung des Gewichts plus/minus 10 kg (was wenig ist). Das Verhältnis von Effektgröße zu Standardabweichung beträgt nun 0,05 (0,5kg/10kg), d.h. der erwartete Effekt liegt bei 5% der Standardabweichung. Die gesuchte Zahl der Probanden für einen entsprechenden Gruppenvergleich beträgt nun 1600, denn die Wurzel aus 1600 ist 40, und 40 mal 0,05 ergibt den kritischen Wert 2.

Natürlich ist eine vollständige Analyse der jeweils erforderten Probandenzahlen komplexer, sie ändert aber nichts an der Größenordnung ihres Verhältnisses. Wenn man der Studie eine andere Fragestellung bzw. ein anderes Design unterschiebt als sie hatte, wird man in der statistischen Bewertung fehllaufen.

Aus den genannten Gründen ist klar, dass erstens das Design experimenteller Studien wie der genannten der Expertise bedarf und zweitens die Kritik an der Fallzahl präpotent wirkt.

(7) Warum wurde die Studie durchgeführt, wo doch z.B. Frampton et al. (2002) teils bei 0,6 ppm und 1,5 ppm Effekte gezeigt haben?

- (a) Die Datenlage zeigt, dass Effekte von NO₂ keineswegs konsistent auftreten und dass sie reproduziert werden sollten, bevor man ihnen traut. Die Arbeit von Frampton et al. (2002) ist gut, zeigt aber keine homogenen Antworten. Vor allem erfolgten keine angemessenen statistischen Vergleiche zwischen Kontrollexposition und Exposition gegenüber 0,6 ppm, wie man sich das gewünscht hätte, um herauszufinden, ob die behaupteten Effekte nicht letztlich alleine auf 1,5 ppm zurückgehen. Man kann sich als Forscher auch fragen, ob die meisten Parameter dieser Studie nicht zu unspezifisch waren, um eine Ankopplung an klinisch relevante Beobachtungen zu erlauben, wie wir es mit unseren Messungen versuchten (siehe Frage 4 b-d). Die Inhomogenität von Studienergebnissen ist nicht nur generell eine Erfahrung in medizinischen Studien, sondern betrifft speziell solche, die nahe der Nachweisgrenze operieren, wie das oft bei Studien zu Luftschadstoffen schon aus ethischen Gründen der Fall ist.
- (b) Wir erwarteten von den modernen, nichtinvasiven, stärker gezielten Methoden *in summa* eher genauere Ergebnisse und stärkere Effekte als von den traditionellen Methoden.
- (c) Ferner - und ganz wesentlich - diente der Einsatz nicht oder wenig invasiver, avancierter Methoden (vor allem exhalierendes bronchiales und alveoläres NO, kombinierte NO/CO-

Diffusionskapazität, auch Nasensekret) dem Zweck, Monitoring-Instrumente zu testen, die in der Praxis, d.h. am Arbeitsplatz und der Epidemiologie tatsächlich einsetzbar sind. Das ist bei der bronchoalveolären Lavage, die Frampton et al. (2002) verwendeten, sicher nicht der Fall. Die Studie hatte also wesentlich auch einen methodologischen Aspekt, der sie wissenschaftlich attraktiv machte. Sie ist in keiner Weise eine Wiederholung der Studie von Frampton et al., und weder wurde sie noch wird sie durch diese überflüssig gemacht.

(8) Warum wurde kein Antrag z.B. bei der DFG gestellt?

- (a) Angesichts des Förderspektrums der DFG wäre der Antrag praktisch aussichtslos und seine Erstellung „wasted time“. Bei anderen potentiellen Förderern (Unfallversicherungen, Ministerien usw.) musste man nach aller Erfahrung damit rechnen, dass jeweils andere Institutionen als zuständig deklariert worden wären, da es nicht zum eigenen Aufgabenbereich gehörte, MAK-Werte zu prüfen.
- (b) Umgekehrt ist eine solche Untersuchung aus „Bordmitteln“ nur noch in Ausnahmefällen zu finanzieren, da im Laufe steter Etat Kürzungen die Mittel immer knapper werden und die Ausfälle durch eingeworbene Drittmittel ersetzt werden müssen. Dieser Prozess wird regelrecht forciert, indem man umgekehrt die Einwerbung von Drittmitteln, gewichtet nach ihrer Herkunft, bei der Vergabe regulärer Mittel als Positivum bewertet.
- (c) In Anbetracht dieser Lage ist man über jeden potentiellen Förderer froh, sofern dieser die wissenschaftliche Unabhängigkeit gewährleistet und auf Design, Auswertung und Publikation keinen Einfluss nimmt.

Diese Antwort zielt darauf ab, der Realität der Forschungsfinanzierung Rechnung zu tragen.

(9) Warum genügen nicht epidemiologische Studien?

Sie tun das nicht in der Arbeitsmedizin und noch weniger in der Umweltmedizin, in der Expositionen noch weniger gut zu definieren sind als am Arbeitsplatz. Soweit - und das gilt nur sehr begrenzt und indirekt (siehe auch Frage 13) - die Studie auch Schlüsse zulässt, die für eine Langzeitexposition bzw. deren Untersuchung relevant sind, ist Folgendes zu bemerken. Das Problem epidemiologischer Studien besteht darin, dass gerade Stickstoffdioxid oft sehr eng mit anderen Schadstoffen korreliert. Das macht die Abtrennung kausaler Effekte schwierig, und dabei helfen auch noch so raffinierte statistische Modelle nur sehr begrenzt, im Gegensatz zu Experimenten. Die Rechenmodelle sind inzwischen oft sehr komplex und mehrstufig, sowie mit multiplen Korrekturen versehen, die alle wiederum mit bestimmten Annahmen einhergehen, so dass sich ab und an sogar scheinbar „schützende“ Effekte einzelner Schadstoffe ergeben. Als Wissenschaftler ist man primär an sorgfältiger Aufklärung und Kausalität, weniger an pauschalen Summationen und bloßen Korrelationen interessiert, und auch für effiziente regulatorische Maßnahmen sollte man sicher differenzieren können.

(10) Wozu braucht man überhaupt experimentelle Studien am Menschen?

Experimentelle Studien sind aus den folgenden Gründen hilfreich.

- (a) Man benötigt sie grundsätzlich, um Mechanismen aufzuklären; hierbei können auch Zellkulturversuche helfen (Holz et al. 1995), die jedoch für eine Festsetzung von Grenzwerten nicht geeignet sind.
- (b) Man kann mit ihnen die Plausibilität epidemiologisch gefundener Korrelationen prüfen. Das ist naturgemäß für eine Wissenschaft, die an kausalem Verständnis interessiert ist, von zentraler Bedeutung.
- (c) Ein illustratives Beispiel ist Ozon, ein Schadstoff, der vermutlich nicht nur für die Gesundheit bedeutsam sein dürfte, sondern die Situation auch besonders schön illustriert. Man kann in 3-4-stündigen Expositionen des Menschen gegenüber 0,2-0,25 ppm Ozon sehr konsistent und klar funktionelle und zelluläre/biochemische Effekte sehen, in der Größenordnung von 0,1 ppm unter extremen Bedingungen ebenfalls (z.B. ca. 6 Stunden Fahrradfahren bei hoher Belastung oder vierfach wiederholte Expositionen plus Allergeninhalation, z.B. Holz et al. 2002). In der Umwelt kann man sicher bei 0,05 ppm und auch darunter akute Effekte auf die Lungenfunktion und zelluläre/biochemische Messgrößen erkennen, die den experimentell beobachteten entsprechen. Auch sieht man langfristige Zusammenhänge mit Asthma und sogar IgE-Spiegeln im Blut, die durchaus plausibel sind. Mithin liegt der Faktor zwischen experimenteller (=kausal nachgewiesener) Wirkung und epidemiologischer Assoziation (=beobachtete Korrelation) bei ca. 3-5, nicht darüber. Das ist in Anbetracht der massiven Unterschiede der Studiensituation hinreichend, um die epidemiologisch beobachteten Effekte völlig glaubwürdig zu machen. Dies gilt auch, weil Ozon zwar oft mit anderen Schadstoffen korreliert, aber negativ und viel weniger als diese untereinander, und daher einigermaßen gut statistisch abtrennbar ist.
- (d) Für NO₂ ist die Diskrepanz zwischen den Konzentrationen, die experimentell konsistent als wirksam aufgezeigt werden können, und den epidemiologisch als wirksam angenommenen viel größer. Das wirft im Vergleich zu Ozon (und auch Feinstaub) für den Wissenschaftler die Frage nach der Kausalität anstelle einer Korrelation (Effekte durch assoziierte Komponenten) auf, vor allem angesichts der verglichen z.B. mit Ozon geringen biologischen Wirksamkeit von NO₂. Unter gut definierten Expositionsbedingungen findet man für NO₂ im Innenraum (Gasöfen, Gasherde) und speziell für die Symptome von Asthma bei Kindern Zusammenhänge, die man eher kausal erklären kann. In der Außenluft sind aber noch viele andere Komponenten.

(11) Lässt die Studie über die MAK-Frage hinaus Rückschlüsse auf die Umwelt-Situation zu?

Ja und nein. Ganz klar nein in regulatorischer Sicht. Dies gilt umso mehr, als man dafür die Gesamtexposition gegenüber Abgasen heranziehen sollte, solange man keine genügend differenzierenden Daten hat. Ja z.B. in dem unter Punkt 13 skizzierten Sinne, der hauptsächlich oder ausschließlich zur Leitung der weiteren, präziseren, kausal orientierten

Forschung von Interesse ist.

(12) Wie ist die Studie wissenschaftlich einzuordnen?

- (a) Bis ca. zum Jahre 2000 wurden mehrere Studien zum NO₂ durchgeführt, aber in zunehmend geringerer Zahl. Das entsprach dem Eindruck vieler Forscher, dass Ozon und Feinstaub erstens gesundheitlich wichtiger und zweitens experimentell „attraktiver“ waren, vor allem auch deshalb, weil man bei NO₂ experimentell „so wenig sah“, was konsistent und interpretierbar war. Insofern war die Studie ein „Nachzügler“, der für einen bestimmten Zweck erfolgte, nämlich die MAK-Fragestellung. Es wäre aber kontraproduktiv gewesen, diese Fragestellung im Manuskript zur Begründung anzuführen. Erfahrungsgemäß heißt es leicht, man solle doch in einer nationalen Zeitschrift veröffentlichen, sobald man die Situation in einem spezifischen Land als Begründung anführt.
- (b) Die Studie wurde bei einer auf dem Felde angesehenen Zeitschrift eingereicht, die weit davon entfernt ist, alle eingereichten Manuskripte anzunehmen. Sie wurde von anonymen Gutachtern bewertet, die nach aller Wahrscheinlichkeit dem Kreis der Experten angehörten, die mit solchen Studien Erfahrung haben und daher zu einem angemessenen Urteil befähigt sind. Die relevante „Scientific Community“ wird von denen gebildet, die auf dem Gebiet tatsächlich arbeiten und Expertise haben. Ethische Probleme oder moralische Affektionen kamen seitens der Gutachter nicht zur Sprache. Ihre Kritik zielte wesentlich auf eine Kürzung des Manuskriptes. Aus solchen Gründen wurden weitere, aufwendig erhobene Daten (elektronische Nase mittels differentieller Ionenmobilitäts-Spektrometrie) weggelassen, die ebenfalls allesamt keinerlei Effekt von NO₂ andeuteten.
- (c) Der wissenschaftliche Erkenntnisprozess verläuft kumulativ. Da man bei der experimentellen Untersuchung von Luftschadstoffen beim Menschen naturgemäß oft nahe der Auflösungsgrenze arbeitet, ist (A) die Reproduktion von Ergebnissen von Bedeutung. Nur was wiederholt konsistent gezeigt wurde, kann man wirklich glauben. Sodann hatte (B) die Studie wesentlich die Aufgabe, mittels moderner, nichtinvasiver Messverfahren (siehe Frage 4) die Wirkung von NO₂ zu überprüfen. Solche Verfahren haben - im Gegensatz zu invasiven, endoskopischen Techniken wie der bronchoalveolären Lavage (BAL) - den unschätzbaren Vorteil, in der Praxis zu Monitoring-Zwecken tatsächlich eingesetzt werden zu können. Zu diesem Thema war für NO₂ nichts auch nur halbwegs Hinreichendes bekannt, schon gar nicht in der gezielten Zusammenstellung der verwandten Methoden.
- (d) Eine Studie nicht zu publizieren, weil „negativ“ und im Widerspruch zu den eigenen Erwartungen oder den Erwartungen anderer, entspricht weder dem individuellen wissenschaftlichen Ethos noch den generellen Erfordernissen einer Forschung, die eine korrekte Beschreibung der Realität zum Ziel hat. Dies gilt umso mehr, wenn die Qualität der Methoden durch langjährige Erfahrung und ausgedehnte methodologische Vorarbeiten zu ihrer Verbesserung gewährleistet ist (siehe Frage 4). Die interne Neigung oder externe Anmutung, „negative“ Studien nicht zu publizieren, erzeugen ein ernsthaftes Problem in der medizinischen Forschung, das unter dem Namen „publication bias“ bekannt ist. Wer über die Jahre oft genug „positive“, aber nicht reproduzierbare

Effekte oder Assoziationen erlebt hat, die sich später ins Fiktive auflösten, wird auch „negativen“ Studien ein großes Gewicht beimessen, falls sie gut gemacht sind. In welche Richtung man aus pathophysiologischer Sicht das Ergebnisse der Arbeit fruchtbar interpretieren könnte, um Hinweise für künftige Studien zu erhalten, ist unter Frage 13 ausgeführt. Somit erfolgte die Arbeit nicht nur formal *lege artis*, sondern erbrachte auch einen Erkenntnisgewinn.

Aus den genannten Gründen ist eine Kritik an der Studie als „nicht innovativ“ oder „qualitativ minderwertig“ fehl am Platze.

(13) Was könnte man als Wissenschaftler aus den Ergebnissen mitnehmen?

- (a) „Negative“ Ergebnisse können hilfreich sein, wenn man die Gesamtlage in Rechnung stellt. Aus pathophysiologischer Sicht erscheinen die epidemiologischen Assoziationen mit NO₂ erklärungsbedürftig, vor allem was die Sterblichkeit angeht. Was sind am Ende die wirklichen Mechanismen, oder sind das Effekte, die in Wirklichkeit auf andere Schadstoffe zurückgehen? Es erscheint aber sinnvoll, zunächst weiter experimentell in Richtung NO₂ und spezieller Patientengruppen zu suchen. Auch wenn die Studie nicht auf diese Frage abzielte, legen ihre Ergebnisse nahe, dass es vermutlich wenig sinnvoll ist, weiter in Richtung von Entzündungseffekten zu suchen; solche treten, wenngleich in abgeschwächter Form, bei NO₂ auch bei Gesunden auf (z.B. Jörres et al. 1995). Dahin gehört die Frage, welche Probanden man eigentlich für diese weitergehende Frage (wohlgemerkt keine MAK-Frage und nicht die Frage dieser Studie) untersuchen sollte. Das sei im Folgenden vorgeführt, um zu demonstrieren, wie man Studien effektiv und produktiv interpretieren kann.
- (b) Die Idee ist, dass das fein abgestimmte Verhältnis von Belüftung und Durchblutung der Lunge gestört wird. Patienten mit COPD haben schlecht belüftete, sauerstoffarme Lungenbereiche, die in der Folge (auf der Basis eines natürlichen Reflexes) auch schlecht durchblutet werden. Auf diese Weise wird verhindert, dass das Blut sozusagen sinnlos in diese Bereiche geleitet wird. Hierbei spielt NO eine wesentliche Rolle, denn es ist im Organismus ein zentraler Regulator der Gefäßweite. Neben NO₂ in der Luft sollte man also NO im Auge haben; interessanterweise finden sich oft auch epidemiologisch Effekte von NO_x (im wesentlichen gleich NO plus NO₂), manchmal sogar separat von NO. Wenn diese Anpassung aufgrund einer globalen Gefäßerweiterung durch eingeatmetes NO, vielleicht auch direkt oder indirekt durch NO₂ gestört würde, würde die Sauerstoffkonzentration im Blut sinken. Dieser Sauerstoffmangel könnte bei Patienten mit stark vorgeschädigtem Herzen klinisch bedeutsame abträgliche Wirkungen haben. Ein an sich gar nicht großer Effekt auf die Lunge würde durch die Folgewirkung auf Herzfunktion bei Vorerkrankung gewissermaßen „verstärkt“. Natürlich wäre die ethische Vertretbarkeit solcher Untersuchungen sehr sorgfältig zu prüfen; man würde am ehesten Patienten mit COPD untersuchen, die herzgesund sind und bei denen dieser „Verstärkermechanismus“ nicht auftreten kann und keine Gefährdung zu erwarten steht. Würde man Effekte auf den Gasaustausch beobachten, müsste man annehmen, dass diese auch bei herzkranken Patienten auftreten und bei diesen in dem skizzierten Sinne verstärkend und wirklich gefährdend wirksam sein könnten. Bei Gesunden wird man keine derartigen Effekte auf den Gasaustausch beobachten, diese sind also für diese Frage als Probanden ungeeignet. Nachdem die Forschung lange in Richtung

Entzündung gesucht hat, würde nun im Anschluss an die Daten der Autoren von Nieding & Wagner (1979), die nie wirklich überprüft wurden, sozusagen das pathophysiologische Fass wieder aufgemacht werden. Das also sind mögliche, inhaltlich produktive Überlegungen. Sie unterstreichen, dass es solider Kenntnisse bedarf, um überhaupt im Ernst Ergebnisse irgendwelcher Art sinnvoll verwerten zu können.

Aus den genannten Gründen ist aus wissenschaftlicher Perspektive die Kritik an der Studie auf keiner Ebene nachvollziehbar, weder methodologisch noch statistisch noch ethisch noch hinsichtlich der Schlussfolgerungen.