UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Statistical methods for the analysis of complex epidemiology data

Mikel Esnaola Acebes

Statistical methods for the analysis of complex epidemiology data

Mikel Esnaola Acebes

TESI PER COMPENDI DE PUBLICACIONS UPC / 2023

DIRECTORS DE LA TESI:

Juan Ramón González Ruiz
Jordi Sunyer Deu,
Department of Statistics and Operations Research

A mis padres, sin ellos no hubiera
empezado esta tesis.

A Tere, sin ella no la hubiera acabado.

Agradecimientos

En primer lugar me gustaría agradecer a mi director Juan Ramón González Ruiz. Esta tesis no hubiera sido posible sin su dirección, ayuda y paciencia.

También quiero agradecer a Jordi Sunyer Deu por estar siempre dispuesto a ayudarme y permitirme formar parte del fantástico equipo de BREATHE. Gracias a él he podido crecer como profesional y como persona.

Gracias a Pedro Delicado por ser un profesor genial y por haberme ayudado tanto durante estos aos.

Gracias a mis padres y a mi hermano por su apoyo incondicional.

Gracias a Tere por aguantar todo este tiempo y seguir creyendo en mí a pesar de todo. Gracias a mis hijos Valentina, Xabi y Beñat porque, junto con su madre, son la principal razón que tengo para levantarme cada mañana.

Gracias a Xavi y Montse por prestarnos su inestimable ayuda siempre que lo hemos necesitado.

Gracias a James por su ayuda con el inglés y por enseñarme tanto con todas nuestras fructíferas conversaciones.

Por último, gracias a todos mis amigos: gracias a todos los compañeros del CREAL/ISGlobal, gracias a los colegas matemáticos del *dinar del mes* y gracias también a mis amigos de Vitoria.

Resumen

La epidemiología es una ciencia relativamente joven pero en constante evolución. Los enormes avances tecnológicos de las últimas décadas han permitido a los epidemiólogos modernos plantearse preguntas científicas cada vez más ambiciosas. Esto ha originado grandes cantidades de datos, lo que a su vez ha llevado a una explosión de nueva metodología estadística. En esta tesis presentamos cuatro artículos que tratan sobre el análisis de datos epidemológicos complejos. En el primer artículo estudiamos la utilidad y validez de dos tests computerizados para cuantificar la Memoria de Trabajo y la Atención. Para ello nos servimos de Redes Bayesianas para inferir la estructura subyacente de interdependencias entre un conjunto de variables que comprenden proxies sociodemográficos y de desarrollo neurológico. Los resultados demuestran que ambos tests tienen buenas propiedades psicométricas y permiten obtener una mejor comprensión de la estructura subyacente de los datos. Los artículos 2 y 3 tratan sobre el estudio de efectos de pequeña magnitud de contaminantes ambientales en el neurodesarrollo de niños y niñas prepúber. En el artículo 2 estudiamos el impacto sobre el neurodesarrollo de diversos contaminantes como por ejemplo el dióxido de nitrógeno o las partículas ultrafinas. Nuestros resultados apuntan a que algunos de estos contaminantes son potencialmente perniciosos para el desarrollo neurológico. En el artículo 3 exploramos más detalladamente el rol de uno de estos contaminantes (la materia particulada 2.5 o PM2.5). Para ello nos valemos de la factorización no negativa de matrices para estimar las fuentes de los diferentes componentes hallados en PM2.5. Tanto en el artículo 2 como en el artículo 3 aplicamos modelos de efectos mixtos con varios efectos aleatorios anidados para tener en cuenta la naturaleza jerárquica y correlacionada de los datos. Por último, en el artículo 4 presentamos nuestro propio método para el análisis de datos de expresión RNA-Seq y, más especificamente, para la detección de genes diferencialmente expresados entre dos o más condiciones. Primero utilizamos ejemplos reales para mostrar por qué la metodología previamente existente basada en las distribuciones de Poisson o Binomial Negativa no es suficientemente flexible para capturar la distribución real de los datos de expresión obtenidos mediante RNA-Seq. A continuación explicamos nuestro método basado en la familia de distribuciones Poisson-Tweedie. Por último demostramos que la flexibilidad de la Poisson-Tweedie permite a nuestro método capturar con mayor precisión las dinámicas de la expresión de dichos datos.

Abstract

Epidemiology is a relatively young and rapidly evolving science. Recent technological breakthroughs have allowed modern epidemiologists to raise increasingly ambitious research hypotheses. This has originated vast amounts of data, which in turn has lead to an explosion of new statistical methodology. In this thesis we present four papers on the analysis of complex epidemiology data. In paper 1 we study the utiliy and validity of two computerized tests for the quantification of Working Memory and Attention. We use Bayesian Networks to learn about the structure of interdependencies between a set of variables comprising sociodemographic and neurodevelopmental proxies. The results show that both tests have good psychometric properties and allow us to have a better understanding of the underlying structure of the data, which can then be incorporated into posterior analyses. Papers 2 and 3 deal with the study of small magnitude effects of environmental pollutants on the neurodevelopment of prepuber children. In paper 2 we study the impact on neurodevelopment of several pollutants such as nitrogen dioxide or ultrafine particles. Our results suggest that these pollutants are potentially harmful for the neurodevelopment. In paper 3 we explore in detail the role of one of these pollutants (particulate matter 2.5 or PM2.5). We use nonnegative matrix factorization to conduct a source apportionment to estimate the levels of the different components present in PM2.5. In both papers we apply mixed effects models with several nested random effects to account for the hierarchical and correlated nature of the data. Finally, in paper 4 we present our own method for the analysis of RNA-Seq data and, more specifically, for the detection of differentially expressed genes across two or more conditions. We first use real data examples to show why the previously existing methods based on Poisson and Negative Binomial distributions are not able to capture the real distribution of expression data obtained via RNA-Seq. We then explain our method, which relies on the Poisson-Tweede family of distributions. We end by showing that our method is able to capture more precisely the dynamics of expression of RNA-Seq data.

Keywords: epidemiology, DAG, Bayesian Networks, Mixed-effects, multilevel, transcriptomics, RNA-seq

Contents

Introduction

Epidemiology is traditionally defined as the study of the distribution and determinants of health-related states or events in specified populations and the application of this study to the control of health problems, Szklo and Nieto (2007). The ancient Greek physician Hippocrates, who is regarded as the father of medicine, is the first person known to have studied the relationship between disease occurrence and environmental effects. Although some outstanding epidemiologic studies were conducted before the 20th century (the most relevant being John Snow's investigations into the causes of a 19th century cholera outbreak in London) it was not until the second half of the 20th century that modern epidemiology, as we know it today, emerged.

During the 1940s, several large-scale epidemiologic studies were launched, some of which had profound influences on health. For example, the Framingham Heart Study, initiated in 1949, importantly contributed to understanding the aetiology of cardiovascular disease, Dawber et al. (1957). Those years saw also the publication of many epidemiologic studies on the effects of tobacco use on health which eventually led to the turning point report, *Smoking and Health*, United-States (1964), the first one to openly denounce the adverse effects of tobacco on health. This was one of the first studies to gain wide public attention. Since that report epidemiologic research increasingly gained public recognition.

The explosion of epidemiologic activity led to disagreements about basic conceptual and methodological points, which in turn lead to a rapid growth in the understanding and synthesis of epidemiologic concepts, Rothman et al. (2008). For instance, the early studies on smoking and lung cancer

were not only important for their important findings, but also because they proved the validity and utility of case-control studies, while controversies regarding their design led to further develop the epidemiologic theory. Other studies, such as the Framingham Heart study, stimulated the development of the most popular modeling method in epidemiology today: multiple logistic regression, Cornfield (1962).

As the complexity of new epidemiological studies grew bigger, so did the need for novel statistical methodology. The surge of epidemiologic activity in the late 20th century, together with the rapid advance in computational technology, lead to an explosion of new statistical techniques. Bayesian statistics, for example, were previously viewed as unfavorable by many statisticians mainly due to the enormously big required computational power. The advent of powerful computers made feasible the use of computationally heavy algorithms such as Markov Chain Monte Carlo (MCMC), Hastings (1970). As a result, since the 1990s, Bayesian statistics have increasingly gained popularity and have even become the *de facto* method in some specific fields such as in variable selection in biomedicine or in small areas analysis.

The technological breakthroughs of the last decades have not been limited to the field of computation. Many other advances have radically affected epidemiology and, in consequence, have also revolutionized statistics. These include, among others, the development and rapid popularization of genotyping and expression microarray technologies, the emergence of improved exposure assessment devices and the advent of Big Data.

The present work consists of a compendium of published papers on statistical methodologies and applications to face some of these arising challenges. These papers deal with some of the main issues to consider in the design and analysis of epidemiologic studies. In the first paper Bayesian Networks are used to study the relevance and relationship of several variables previous to their analysis and inclusion in models. In the next two papers, multilevel models are applied to account for the hierarchical structure of the data. Finally, paper 4 deals whith the analysis of transcriptomic data as a mean to improve the characterization of genetic predisposition and to offer a more refined way to identify differentially expressed genes. We will now proceed to separately introduce each of these four papers.

Bayesian Networks

The advances in the field of epidemiology have allowed epidemiologists to ask more complex and ambitious research questions. This, naturally, has led to richer science but it has also posed many new technical difficulties. One of these problems is the need to establish *a priori* the interdependencies between the studied variables in order to discern all possible causating/confounding/effect-modifying/mediating effects.

Most statistical analyses in epidemiologic studies are organized around three different sets of variables: exposures, outcomes and confounders. While exposures and outcomes are normally determined by the research hipothesis under investigation, confounders are not so clearly defined and need to be first identified in order to account for them in the subsequent analyses. Causal diagrams known as Directed Acyclic Graphs (DAGs) have been proposed to study the relationship between these sets of variables, Greenland et al. (1999). DAGs are a useful way of representing biases other than confounding, such as selection and recall bias. Their main advantage is that, while being visual and intuitive, they are also mathematically rigorous, Hernán et al. (2002).

While DAGs are useful as a graphical representation of the inter-relationships of a set of variables, they do not naturally incorporate uncertainty. The relationships between variables described by DAGs are not established in terms of probabilistic distributions. As a result, DAGs are mainly useful when summarizing one's qualitative beliefs about the casual structure but cannot be directly applied to learn the casual structure of a set of variables from observational data. In order to be able to do this DAGs must be refined by adding probabilistic dependencies between connected nodes.

A Bayesian Network (BN, Pearl (1985)) is a graphical model that compactly describes the dependency structure between a given set of variables. It consists of a DAG that encodes the conditional dependencies between the variables and a set of local probability distributions associated to each of the variables. Nodes in DAGs represent variables ($\mathbf{X} = (X_1, \ldots, X_n)$) and arrows represent conditional dependence assertions. We will say that a node X_j is a parent of another node X_i if there is a directed edge from X_j to X_i. The local independencies of the graph can be interpreted as following: each node X_i is conditionally independent of its nondescendants given its parents. More formally,

$$\forall X_i \, (X_i \perp \{X_j : X_i \notin \Pi_j\} | \Pi_i)$$

where Π_i represents the set of parent nodes of X_i.

As a result, the local probability of a node conditioned on the rest of the nodes can be explicited as

$$P(X_i|(\mathbf{X}\backslash X_i)) = P(X_i|\Pi_i)$$

Applying the chain rule of probability we obtain the expression of the joint probability distribution for the whole Bayesian Network:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i|\Pi_i)$$

A BN can accommodate both discrete and continuous variables and different probability distributions for each of them. However, in most cases the inference becomes intractable if we do not restrict to certain distributions. The most common approach, and the one we will use in our present work, is to consider all variables to be discrete. Therefore, nodes can be modeled with multinomial distributions and Dirichlet (multivariate generalization of the beta distribution) priors.

A Bayesian measure of the goodness of fit of a Bayesian Network G is its posterior probability given the observed data D. Using Bayes' Theorem:

$$P(G|D) = \frac{P(D|G)P(G)}{\sum_G P(D|G)P(G)}$$

The problem of searching for the Bayesian Network that best fits our data is that the number of possible graphs grows super-exponentially with the number of variables, Chickering et al. (2004). As a result, it is practically impossible to compute exactly this posterior probability. Therefore, instead of the exact expression of the posterior probability we will use a proportional one:

$$P(G|D) \propto P(D|G)P(G) = P(D, G)$$

This leads to a closed form for the scoring function of a BN. But the search of the highest scoring graph given our data is far from trivial. Even for small domains, it is unfeasible to list all possible BNs with their corresponding score. Therefore, a convenient search method that ensures pseudo-optimality of the obtained BN is needed. In our study, to gain validity, we have applied three different search methods: greedy search, structure MCMC and order MCMC.

The greedy search method, also known as hill climbing, is a heuristic method that is widely used due to its speed and simplicity. However, it rarely yields good results because it easily gets trapped in local optima. The method starts with a user prespecified BN and it looks for the perturbation that most improves its score. Perturbations are normally defined as deleting, adding or reversing an edge. This is repeated until a graph with no possible improving perturbation is found, i.e., until a local optima is found. There is no possible way to ensure the optimality or pseudo-optimality of the resulting Bayesian Network. Nevertheless, we decided to include this method in our analyses for illustrative purposes only.

Structure MCMC is based on a Metropolis-Hastings random-walk over the space of possible graphs. In each iteration a perturbation of the graph (adding, deleting or reversing an edge) is randomly chosen and the score of the perturbed graph is updated. The main difference with the greedy search is that this method is able to escape from local optima, as it sometimes accepts updated graphs with worse scores. The problem is that, given the vast space of possible graphs, it could take a large number of iterations to converge, i.e., to reach an optimal or pseudo-optimal BN. Though there is no proper way to assess the optimality of the obtained solutions, one can parallely run separate instances (normally called chains) of the structure MCMC with different starting graphs. If the resulting BNs are similar or equal for all chains (usually denoted as chain mixing) we will say that they converged. This, of course, is not a foolprof guarantee of the optimality of the solution but, if done with a sufficiently large number of chains, it is normally enough to ensure the robustness and reliability of the results.

Order MCMC is also based on a Metropolis-Hastings random-walk but, instead of traversing the space of possible graphs, it searches in the order space. By order space we refer to the topological order of graphs based on the parenting of its nodes. We can define $\prec$ as a total order and will say that a graph G is consistent with $\prec$ if

$$X_i \in \Pi_j \text{ in G } \implies i \prec j$$

Obviously, there are many possible graphs consistent with a given order and viceversa: there can be many different consistent orders for a certain graph. As a result, searching for the best order given some data is not enough, as many graphs will be consistent with it. But, as the order space is significantly smaller than the graph space, the problem of searching for an optimal or pseudo-optimal Bayesian Network can be simplified by first

looking for the best order and afterwards looking for the best graph that is consistent with the obtained order. Though, as in the case of structure MCMC, there is no way to assess the optimality of the obtained solution, running separate instances of the algorithm and checking the similarity between the obtained BNs is a good way to ensure convergence of the method.

Summarizing, BNs can be applied to observational data with the aim of finding the casual structure that best describes the interdependencies between variables. This turns them in an immensely useful tool for epidemiologists as a mean of understanding the complex structure of the observed data. This knowledge can then be incorporated into subsequent analyses.

In the first presented paper we apply BNs to learn about the complex structure of interdependencies between a set of variables comprising socio-demographic and neurodevelopmental proxies as part of the BREATHE project.

Mixed-effects models

In recent years epidemiological studies have become increasingly more complex. This can be attributed to factors that include modern technological advances, the emergence of macrostudies and even multicenter studies possibilitated by multinational financing institutions such as the National Institute of Health (NIH) or the European Research Council (ERC), more ambitious research questions such as the study of small magnitude environmental/genetic effects or the concurrent study of individual and contextual (groupal) exposures. As a consequence, new statistical methodology has emerged to cope with these new complexities.

Historically, the emphasis of epidemiology has suffered a profound shift from an ecological to an individual perspective. In the 19th century, public health was essentially ecological, i.e., it focused on relating the environmental and community characteristics to health and disease. But later, during the 20th century, as the importance of chronic diseases grew, emphasis shifted to individual-level factors. As a result, the study of the causes of disease changed from the environment as a whole to specific factors and individual behaviors. This change of paradigm was accompanied by a progressive individualization of risk, which in turn helped to support the idea that risk is individually determined rather than socially determined, Diez-Roux (1998).

The late 20th century saw a growth of literature on the potential impor-

tance of ecological variables such as area-based measures of socioeconomic status. As a consequence, several renowned epidemiologists advocated for a new paradigm aiming to integrate these different levels, thus leading to a more public health-oriented epidemiology. This new approach facilitated the popularization of already existing statistical methodology especially designed to deal with the hierarchical/multilevel nature of the produced data. One of these methodologies was Mixed-effects modeling, also known as Multilevel modeling.

Mixed-effects models (and their extension, Generalized Mixed-effects models) extend linear models by allowing a wide variety of correlation patterns or variance-covariance structures to be explicitly modeled, Pinheiro and Bates (2000). The standard linear model assumes independently sampled observations by considering a unique random effect, i.e., the error term. As a result, it is not appropriate when modeling hierarchical data (data collected via sampling at two or more levels, one nested within the other) or longitudinal data (repeated measurements). Mixed-effects models allow to take account of dependencies in hierarchical, longitudinal and other dependent data by including more than one source of random variation, i.e., more than one random effect.

In its simplest form the linear mixed-effects (LME) model can be expressed as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

where $\mathbf{y}_i$ represents the response vector for the i-th group, $\boldsymbol{\beta}$ is the vector of fixed effects, $\mathbf{b}_i$ is the vector of random effects, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are fixed-effects and random-effects regressor matrices respectively and $\boldsymbol{\epsilon}_i$ is the within-group error vector. This model is useful when the data presents a single level of grouping.

The formulation for single-level LME models can be easily extended to incorporate multiple, nested levels of random-effects. This is of special interest when the data to be modelled is organized hierarchically. While the single-level LME models only allow a unique level of grouping, the multilevel LME models can accommodate as many nested grouping levels as desired. In our case, the experimental design of the BREATHE project was based on 2 nested levels of grouping: individuals within schools and several repeated measurements within each individual*. The multilevel LME

*For more information about the BREATHE project refer to the **Introduction to the BREATHE project** in page 85.

with two nested random effects can then be formulated as

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{i,j}\mathbf{b}_i + \mathbf{Z}_{ij}\mathbf{b}_{ij} + \boldsymbol{\epsilon}_{ij}$$

where $\mathbf{y}_{ij}$ are the response vectors at the innermost level of grouping (i refers to the first-level grouping and j to the second-level grouping within firs-level group i), $\mathbf{X}_{ij}$ are the fixed-effects model matrices, $\mathbf{b}_i$ are the first-level random effects, $\mathbf{b}_{ij}$ are the second-level random effects with corresponding model matrices $\mathbf{Z}_{i,j}$ and $\mathbf{Z}_{ij}$, and $\boldsymbol{\epsilon}_{ij}$ are the within-group errors.

Papers 2 and 3 deal with the study of small magnitude effects of environmental pollutants on the neurodevelopment of prepuber children as part of the BREATHE project. This project relies on a nested cohort design where the outcome variables have repeated measurements and the explanatory variables can be individual or groupal. Moreover, while most of the outcome variables are normally distributed, some of them are counts and need to be modelled using a Poisson or Negative Binomial distribution. To cope with all these difficulties, we decided to use Mixed-effects models. In Paper 2, we applied this modelling technique to explore the association of several pollutants with different neurodevelopment proxies. In Paper 3, we further explored the role of one of this pollutants (PM2.5) by first categorizing it into its possible different sources and then applying Mixed-effects modeling to individually study the association of each of them with neurodevelopmental deceleration.

Differential expression

One of the biggest revolutions in, not only epidemiology, but in the whole biomedical sciences has been the advent of OMICS. The word OMICS refers to many fields of study in biological sciences, all of them ending with the *-omics* suffix. These include, among others, genomics (the study of the structure, function, evolution and mapping of genomes), transcriptomics (the study of gene expression), proteomics (the study of proteins) and epigenetics (the study of changes in organisms caused by modification of gene expression rather than the alteration of the genetic code itself).

Since their inception, OMICS data have played a major role in epidemiologic studies. The first available OMICS data were derived in the 1970s using Sanger sequencing, Sanger et al. (1977). This technology allows to determine the nucleotide sequences in DNA. The availability of DNA sequencing allowed epidemiologic researchers to add and include genetic information in their studies. Despite being a revolutionary technique, Sanger

sequencing is slow and very expensive. As a result, it was not normally feasible to use it on large scale epidemiologic studies.

Some years later, in the 1990s, microarrays were developed. A microarray is a two-dimensional array normally used to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome, although it can also be used with other types of OMICS data such as proteomics. The processing of microarrays is significantly cheaper and faster than that of Sanger sequencing and, as a result, this technology rapidly became popular in epidemiologic studies. Unfortunately, microarrays also present some serious limitations; mainly that they heavily rely upon existing knowledge about the genome sequence, as each microarray can only provide information about the genes that are included in the array, Wang et al. (2009).

At the onset of the 21st century a new set of highly efficient, rapid and low cost DNA sequencing platforms arose commonly named Next Generation Sequencing (NGS). NGS relies on massively parallel sequencing of millions of small fragments of DNA. While Sanger sequencing required over a decade to sequence an entire human genome, using NGS this could be achieved within a single day. Furthermore, not only can NGS be used to obtain the whole genome sequence of an individual (or organism) but it can also be applied to the transcriptome (including mRNAs, non-coding RNAs and small RNAs). In particular, RNA-seq refers to the profiling of RNA sequences using NGS technology. It is the first sequencing-based method that allows to survey the entire transcriptome in a high-throughput and quantitative manner.

The study of RNA-seq data is of particular interest in the field of epidemiology. RNA-seq measures the expression level of genomic loci (specific position on a chromosome where a particular gene or genetic marker is located). Thus, a natural application of RNA-seq in epidemiology is the study of differentially expressed regions between two or more groups (such as, for instance, the study of differentially expressed genes between healthy and cancerous cells).

RNA-seq offers a significant improvement over previous transcriptome surveying methods such as microarrays but it also poses many new statistical challenges. One of the main challenges is related to the probability distribution of the produced data. Microarrays are measured in signaling terms and, therefore, can be modeled using a normal distribution, Ritchie et al. (2015). On the other hand, RNA-seq produces counts which need to be modelled using an appropriate discrete distribution. The first proposed methods were

based on the Poisson distribution but they soon proved to be inadequate due to the inability of this distribution to cope with some oddities presented by RNA-seq data (mainly overdispersion and zero-inflation). Later, methods based on the Negative-Binomial distribution were developed but, although these methods proved to be significantly better than those based on the Poisson distribution, they still lacked flexibility to cope with the complexities of RNA-seq data.

Two of the most widely used methods for the analysis of RNA-seq data are *edgeR* (together with *voom*) and *DESeq* (and its successor *DESeq2*). *edgeR*, Robinson et al. (2010), is based on the Negative Binomial distribution assuming that mean and variance are related by $\sigma^2 = \mu + \alpha\mu^2$, where α is a proportionality constant estimated from the data that remains the same throughout the experiment. This mean-variance relationship is too simplistic and thus is not flexible enough for capturing the wide dynamic range of RNA-seq count data. Later, the authors of *edgeR* (who also happen to be the authors of *limma*, one of the most popular methods for the analysis of expression microarrays) presented *voom*, a method for better estimating the mean-variance relationship based on their previous microarray method *limma*.

DESeq, Anders and Huber (2010), extended the *edgeR* method by allowing a more general, data-driven relationships of variance and mean. For each gene i under condition $\rho(j)$ they proposed a mean-variance relationship that can be expressed as $\sigma^2_{ij} = \mu_{ij} + s^2_j \nu_{i,\rho(j)}$ where $\nu_{i,\rho(j)}$ is a smooth function of a condition-dependent per-gene value and the experimental condition of the sample. This mean-variance relationship is more flexible that the one used by *edgeR*, thus giving better results. Later, the authors presented an updated version of the method called *DESeq2*, Love et al. (2014), in which they added shrinkage estimation for dispersions and fold changes to improve stability and interpretability of estimates.

In Paper 4 we present our own method for the analysis of RNA-seq data based on the Poisson-Tweedie (PT) family of distributions. This family of distributions unify several over-dispersed count data distributions such as the Poisson, Negative-Binomial, Poisson Inverse Gaussian or Pólya - Aeppli, among others. More specifically, the Poisson-Tweedie distribution can be defined from its probability generating function:

$$G_Y(y|a,b,c) = \exp\left\{\frac{b}{a}\left((1-c)^a - (1-cy)^a\right)\right\}$$

when $a \neq 0$, while when $a = 0$ its probability generating function can be

expressed as

$$\lim_{a \to 0} G_Y(y|a,b,c) = \left[\frac{(1-c)}{(1-cy)}\right]^b$$

The parameters (a, b, c) can be reparameterized into (μ, ϕ, a), where μ denotes de mean, $\phi = \sigma^2/\mu$ is the dispersion index and a is the shape parameter. a can be used to define specific count-data distributions that are particular cases of the PT family of distributions, the most relevant of them being:

Value of a	Distribution
$a = 1$	Poisson
$a = 0$	Negative-Binomial
$a = \frac{1}{2}$	Poisson inverse Gaussian
$a = -1$	Pólya-Aeppli
$a \to -\infty$	Neyman type A

The use of the PT family of distributions instead of the Poisson or NB distributions allows to better capture the wide diversity of expression profiles arising from extensively replicated RNA-seq experiments. For example, a RNA-seq expression profile with a wide dynamic range will most likely lead to a heavy tail in the distribution. In such a case, PIG has a heavier tail than NB and this would make it more appropriate. Another similar scenario in which the PT outperforms the Poisson or NB distributions is the existence of (possibly extreme) zero-inflation.

Another advantage of using the PT distribution is its mean-variance relationship which, using the parameterization of Kokonendji et al. (2004), can be expresssed as

$$\sigma^2 = \mu(1 + \mu^{p-1} \exp\{(2-p)\phi_p\})$$

where p denotes the shape parameter of that specific parameterization. It follows that, whereas the NB distribution is only capable of capturing a quadratic mean-variance relationship, the PT family is able to generalize this relationship to any order. As a result, the PT is more convenient when dealing with count data with variable overdispersion.

In paper 4 we show that the PT distribution can outperform the Poisson or NB distributions using both real and simulated data. Our method is also

openly available as an R package in the Bioconductor repositories (package `tweeeDEseq`).

Copy of published papers

I. The n-back test and the Attentional Network Task as measures of child neuropsychological development in epidemiological studies

II. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study

PLOS | MEDICINE

RESEARCH ARTICLE

Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary School Children: A Prospective Cohort Study

Jordi Sunyer*[1,2,3,4], Mikel Esnaola[1,2,3], Mar Alvarez-Pedrerol[1,2,3], Joan Forns[1,2,3], Ioar Rivas[1,2,3,5], Mònica López-Vicente[1,2,3], Elisabet Suades-González[1,2,3,6], Maria Foraster[1,2,3], Raquel Garcia-Esteban[1,2,3], Xavier Basagaña[1,2,3], Mar Viana[5], Marta Cirach[1,2,3], Teresa Moreno[5], Andrés Alastuey[5], Núria Sebastian-Galles[2], Mark Nieuwenhuijsen[1,2,3], Xavier Querol[5]

1 Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Catalonia, Spain, **2** Pompeu Fabra University, Barcelona, Catalonia, Spain, **3** Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain, **4** Institut Hospital del Mar d'Investigacions Mèdiques–Parc de Salut Mar, Barcelona, Catalonia, Spain, **5** Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Catalonia, Spain, **6** Learning Disabilities Unit (UTAE), Neuropediatrics Department, Hospital Sant Joan de Déu, Universitat de Barcelona, Barcelona, Spain

* jsunyer@creal.cat

Citation: Sunyer J, Esnaola M, Alvarez-Pedrerol M, Forns J, Rivas I, López-Vicente M, et al. (2015) Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary School Children: A Prospective Cohort Study. PLoS Med 12(3): e1001792. doi:10.1371/journal.pmed.1001792

Academic Editor: Bruce P. Lanphear, Simon Fraser University, CANADA

Received: September 16, 2014

Accepted: January 9, 2015

Published: March 3, 2015

Data Availability Statement: Data are from the BREATHE study whose authors may be contacted at CREAL (http://www.creal.cat/projectebreathe/).

Funding: The research leading to these results has received funding from the European Research Council under the ERC Grant Agreement number 268479 – the BREATHE project. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

Air pollution is a suspected developmental neurotoxicant. Many schools are located in close proximity to busy roads, and traffic air pollution peaks when children are at school. We aimed to assess whether exposure of children in primary school to traffic-related air pollutants is associated with impaired cognitive development.

Methods and Findings

We conducted a prospective study of children (n = 2,715, aged 7 to 10 y) from 39 schools in Barcelona (Catalonia, Spain) exposed to high and low traffic-related air pollution, paired by school socioeconomic index; children were tested four times (i.e., to assess the 12-mo developmental trajectories) via computerized tests (n = 10,112). Chronic traffic air pollution (elemental carbon [EC], nitrogen dioxide [NO2], and ultrafine particle number [UFP; 10–700 nm]) was measured twice during 1-wk campaigns both in the courtyard (outdoor) and inside the classroom (indoor) simultaneously in each school pair. Cognitive development was assessed with the n-back and the attentional network tests, in particular, working memory (two-back detectability), superior working memory (three-back detectability), and inattentiveness (hit reaction time standard error). Linear mixed effects models were adjusted for age, sex, maternal education, socioeconomic status, and air pollution exposure at home.

Children from highly polluted schools had a smaller growth in cognitive development than children from the paired lowly polluted schools, both in crude and adjusted models

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: ADHD, Attention deficit hyperactivity disorder; ANT, attentional network test; BC, black carbon; EC, elemental carbon; HRT-SE, hit reaction time standard error; NDVI, Normalized Difference Vegetation Index; PM2.5, particulate matter < 2.5 μm; SDQ, Strengths and Difficulties Questionnaire; UFP, ultrafine particle number.

(e.g., 7.4% [95% CI 5.6%–8.8%] versus 11.5% [95% CI 8.9%–12.5%] improvement in working memory, $p = 0.0024$). Cogently, children attending schools with higher levels of EC, NO2, and UFP both indoors and outdoors experienced substantially smaller growth in all the cognitive measurements; for example, a change from the first to the fourth quartile in indoor EC reduced the gain in working memory by 13.0% (95% CI 4.2%–23.1%). Residual confounding for social class could not be discarded completely; however, the associations remained in stratified analyses (e.g., for type of school or high-/low-polluted area) and after additional adjustments (e.g., for commuting, educational quality, or smoking at home), contradicting a potential residual confounding explanation.

Conclusions

Children attending schools with higher traffic-related air pollution had a smaller improvement in cognitive development.

Introduction

Air pollution is a suspected developmental neurotoxicant [1]. In animals, inhalation of diesel exhaust and ultrafine particles results in elevated cytokine expression and oxidative stress in the brain [2,3] and altered animal behavior [4,5]. In children, exposure to traffic-related air pollutants during pregnancy or infancy, when the brain neocortex rapidly develops, has been related to cognitive delays [6–8].

Children spend a large proportion of their day at school, including the period when daily traffic pollution peaks. Many schools are located in close proximity to busy roads, which increases the level of traffic-related air pollution in schools and impairs children's respiratory health [9]. There is currently very little evidence on the role of traffic-related pollution in schools on cognitive function [10]. Though the brain develops steadily during prenatal and early postnatal periods, resulting in the most vulnerable window [1], high cognitive executive functions essential for learning [11] develop significantly from 6 to 10 y of age [12]. The brain regions related to executive functions such as working memory and attention—largely the prefrontal cortex and the striatum [13]—have shown inflammatory responses after traffic-related air pollution exposure [2,14]. We aimed to assess the relationship between long-term exposure to traffic-related air pollutants at school and cognitive development measurements in primary school children within the BREATHE (Brain Development and Air Pollution Ultrafine Particles in School Children) project.

Methods

Funding

The research leading to these results received funding from the European Research Council under ERC Grant Agreement number 268479 for the BREATHE project.

Design

Forty schools in Barcelona (Catalonia, Spain) were selected based on modeled traffic-related nitrogen dioxide (NO_2) values [15]. Low- and high-NO_2 schools were paired by socioeconomic vulnerability index and type of school (i.e., public/private). A total of 39 schools agreed to participate

and were included in the study (Fig. 1). Participating schools were similar to the remaining schools in Barcelona in terms of socioeconomic vulnerability index (0.46 versus 0.50, Kruskal-Wallis test, $p = 0.57$) and NO_2 levels (51.5 versus 50.9 $\mu g/m^3$, $p = 0.72$).

All school children ($n = 5{,}019$) without special needs in grades 2 through 4 (7–10 y of age) were invited to participate, and families of 2,897 (59%) children agreed. All children had been in the school for more than 6 mo (and 98% more than 1 y) before the beginning of the study. All parents or guardians signed the informed consent form approved by the Clinical Research Ethical Committee (No. 2010/41221/I) of the Institut Hospital del Mar d'Investigacions Mèdiques–Parc de Salut Mar, Barcelona, Spain.

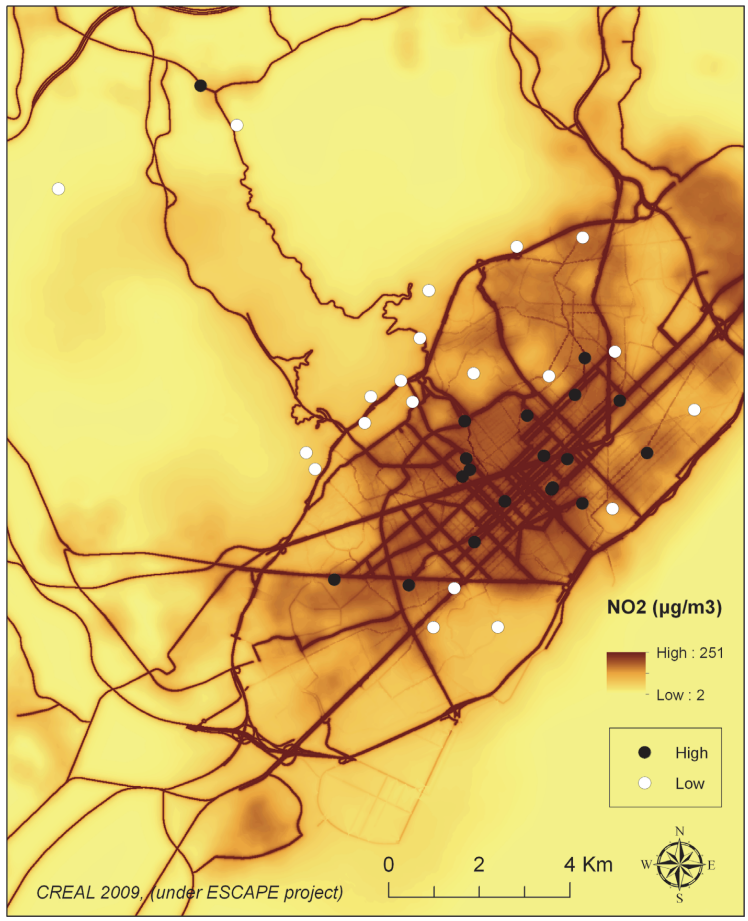

Fig 1. Map of Barcelona and the schools by high or low air pollution by design. Black dots indicate the locations of schools with high air pollution, and white dots indicate the locations of schools with low air pollution, based on NO_2 levels.

doi:10.1371/journal.pmed.1001792.g001

Outcomes: Cognitive Development

Cognitive development was assessed through long-term change in working memory and attention. From January 2012 to March 2013, children were evaluated every 3 mo over four repeated visits, using computerized tests in series lasting approximately 40 min in length. We selected working memory and attention functions because they grow steadily during preadolescence [12,16]. The computerized tests chosen (the *n*-back task on working memory [12] and the attentional network test [ANT] [17]) have been validated with brain imaging [13,17] and in the general population [18]. Groups of 10–20 children were assessed together, wearing ear protectors, and were supervised by one trained examiner per 3–4 children. For the *n*-back test, we examined different *n*-back loads (up to three back) and stimuli (colors, numbers, letters, and words). For analysis here, we selected two-back and three-back loads for number and word stimuli as they showed a clear age-dependent slope in the four measurements and had little learning effect. Numbers and words activate different brain areas. The two-back test predicts general mental abilities (hereafter called working memory), while the three-back test also predicts superior functions such as fluid intelligence (hereafter called superior working memory) [19]. All sets of *n*-back tests started with colors as a training phase to ensure the participant's understanding. The *n*-back parameter analyzed was *d* prime (d'), a measure of detection subtracting the normalized false alarm rate from the hit rate: $(Z_{\text{hit rate}} - Z_{\text{false alarm rate}}) \times 100$. A higher d' indicates more accurate test performance. Among the ANT measures, we chose hit reaction time standard error (HRT-SE) (standard error of reaction time for correct responses)—a measure of response speed consistency throughout the test [20]—since it showed very little learning effect and the clearest growth during the 1-y study period among all the ANT measurements. A higher HRT-SE indicates highly variable reactions related to inattentiveness.

Exposures: Direct Measurements of Traffic-Related School Air Pollution

Each pair of schools was measured simultaneously twice during 1-wk periods separated by 6 mo, in the warm and cold periods of the year 2012. Indoor air in a single classroom and outdoor air in the courtyard were measured simultaneously. The pollutants measured during class time in schools were real-time concentrations of black carbon (BC) and ultrafine particle number (UFP; 10–700 nm in this study) concentration, measured using the MicroAeth AE51 (AethLabs) and DiSCmini (Matter Aerosol) meters, respectively, and 8-h (09:00 to 17:00 h) particulate matter < 2.5 μm ($PM_{2.5}$) measured using a high-volume sampler (MCV). Details of $PM_{2.5}$ filter chemical analysis are described elsewhere [21]. Given the high correlation between continuous BC and elemental carbon (EC) in $PM_{2.5}$ filters ($r = 0.95$), only EC was considered here. Weekday NO_2 was measured with one passive tube (Gradko). We selected EC, NO_2, and UFP given their relation to road traffic emissions in Barcelona, particularly EC [21,22]. In contrast, school's $PM_{2.5}$ was poorly related to traffic because of the relevance of specific school sources in our study [21,23] and was not included here.

Outdoor and indoor long-term school air pollution levels were obtained by averaging the two 1-wk measures. To achieve a better spatial long-term average, EC and NO_2 were also adjusted for temporal variability. Seasonalized levels were obtained by multiplying the daily concentration at each school by the ratio of annual average to the same day concentration at a fixed air quality background monitoring station in Barcelona, operationed continuously throughout the year, as detailed elsewhere [23]. Seasonalized measures had a stronger correlation between the first and the second campaign than non-seasonalized measures (e.g., $r = 0.73$ versus 0.61 for indoor EC and $r = 0.64$ versus 0.62 for indoor NO_2). In contrast, seasonalized UFP had a poorer correlation between the two measurement campaigns than non-seasonalized UFP ($r = 0.38$ versus 0.70 for outdoor UFP and $r = 0.17$ versus 0.40 for indoor levels).

Therefore, non-seasonalized UFP was selected in this study. The correlations between the temporally adjusted annual concentrations of EC and NO_2 at each school and the land use regression annual estimate of BC at each school were 0.73 and 0.74, respectively, indicating good capture of the long-term average concentrations at these schools.

Contextual and Individual Covariates

Socio-demographic factors were measured using a neighborhood socioeconomic vulnerability index (based on level of education, unemployment, and occupation in each census tract, the finest spatial census unit, with median area of 0.08 km^2) [24] according to both the school and home address, as well as through parents' responses to the BREATHE questionnaire on family origin, gestational age and weight, breastfeeding, parental education, occupation, marital status, smoking during pregnancy, environmental tobacco smoke at home, commuting mode, and use of computer games. Standard measurements of height and weight were performed to define overweight and obesity [25]. Attention deficit hyperactivity disorder (ADHD) symptoms (ADHD/DSM-IV Scales, American Psychiatric Association 2002) were reported by teachers. Parents completed the Strengths and Difficulties Questionnaire (SDQ) on child behavioral problems [26].

Noise in the classroom before children arrived to school (hereafter called noise) was measured as the best marker of traffic noise exposure and was included here as a covariate. Data were obtained from comprehensive noise measurements conducted during the second 1-wk campaign of air pollution sampling. Three consecutive 10-min measurements of equivalent sound pressure levels (in A-weighted decibels) at different distributed locations within the classroom were performed over two consecutive days using a calibrated SC-160 sound level meter (CESVA; ±1.0 dB tolerance [type 2], range: 30–137 dB). As we aimed to register traffic and background noise levels, any unusual sounds were deleted, and measurements were conducted before children arrived to school (before 9:00 A.M.). For robustness, we averaged the 30-min measurements from the two consecutive days, though they showed high reproducibility. Short-term noise measurements as short as 5 min have been shown to represent long-term averages [27].

Exposure at home to NO_2 and BC ($PM_{2.5}$ absorbance) at the time of the study was estimated at the geocoded postal address of each participant using land use regression models, details of which are explained elsewhere [15]. Similarly, school and residential surrounding greenness was measured in buffers of 100 m around the address based on the Normalized Difference Vegetation Index (NDVI) derived from Landsat 5 Thematic Mapper data. Residential history was reported by parents. The longest held address was used in 174 children (5.9%) who lived in two homes over the study period. Distance from home to school was estimated based at the geocoded postal address of each participant and school.

Statistical Analysis

A total of 2,715 (93.7%) children with complete data (i.e., repeated outcome at least twice and individual data on maternal education and age) were included. They performed 10,112 (93.1%) tests. Because of the multilevel nature of the data (i.e., visits within children within schools), we used linear mixed effects models with the cognitive parameters (test performance) from the four repeated visits as outcomes and random effects for child and school. Age (centerd at visit 1) was included in the model in order to capture the growth trajectory of cognitive test performance. An interaction between age at each visit and school air pollution was included to capture changes in growth trajectory associated with school air pollution exposure. The main effect of air pollution (AP), which was also included in the model, captures the baseline (visit

1) differences in cognitive function that are associated with air pollution (model 1):

$$Y_{sit} = \beta_0 + \beta_1(\text{Age}_t - \text{Age}_1) + \beta_2\text{AP} + \beta_3(\text{Age}_t - \text{Age}_1)\text{AP} + u_s + v_{i(s)} + \varepsilon_{sit} \quad (1)$$

where Y_{sit} is the cognitive test result for subject i in school s at visit t, $t = \{1,2,3,4\}$; u_s is random effects at school level, assumed to be normally distributed with mean 0 and variance σ_u^2; $v_{i(s)}$ is random effects associated with subject i in school s, assumed to be normally distributed with mean 0 and variance σ_v^2; and ε_{sit} is the model residuals, assumed to be normally distributed with mean 0 and variance σ_ε^2.

This model was further adjusted for potential confounders selected with directed acyclic graphs. Based on all socio-demographic and contextual covariables mentioned above, we used the program DAGitty 2.0 [28], with a priori definition of the temporal direction of the events, to draw causal diagrams. The final adjusted model (model 2) included additional coefficients for sex, maternal education (less than/primary/secondary/university), residential neighborhood socioeconomic status, and air pollution exposure at home:

$$\begin{aligned} Y_{sit} = {} & \beta_0 + \beta_1(\text{Age}_t - \text{Age}_1) + \beta_2\text{AP} + \beta_3(\text{Age}_t - \text{Age}_1)\text{AP} + \beta_4\text{Sex} \\ & + \beta_5\text{Mat\_educ\_primary} + \beta_6\text{Mat\_educ\_secondary} + \beta_7\text{Mat\_educ\_university} \\ & + \beta_8\text{Neighborhood\_socioeconomic\_status} + \beta_9\text{Air\_pollution\_exposure\_at\_home} + u_s \\ & + v_{i(s)} + \varepsilon_{sit} \end{aligned} \quad (2)$$

The interactions between age and maternal education and socioeconomic status were unrelated to cognitive development ($p = 0.33$) and were not included in the models. Other variables such as quality of the test (i.e., room density and noise) and hour, day of the week, temperature, and humidity at test performance were not included in the final model after assessing their inclusion in the multivariate model and obtaining no change in the school air pollution coefficient (i.e., <1%).

School air pollution exposure was first treated as a dichotomous variable based on the high/low air pollution classification of schools used in the design stage. In a second step, we fitted the same models but replaced the binary air pollution variable by the direct measurements of air pollution levels either inside or outside the schools as quantitative exposures. Linearity of the relation between air pollution and cognitive tests was assumed since using multiple polynomial models did not improve model fit. Furthermore, to assess whether a part of our observed associations was due to potential residual confounding, models were adjusted for all covariates referred to above, both individual (e.g., smoking at home or commuting [distance and walking mode]) and contextual (e.g., greenness or noise). Sensitivity analyses were also conducted to assess effect modification by high-/low-air-pollution school, type of school, and residential neighborhood socioeconomic status in order to explore the potential for residual confounding, and by sex, maternal education, ADHD symptoms, and obesity in order to assess susceptibility. Both stratified analyses and modeling of the third-order interaction term with age, air pollution, and the third variable in the regression models were conducted.

Sample size was calculated based on a previous study that showed differences in executive function (mean 100, standard deviation 15) of four points by carbon particle interquartile range [6]. One would need 800 individuals to detect a difference of four points between the first and last categories of air pollution exposure (assuming exposure is divided into four groups according to quartiles) with a statistical power of 80% and alpha = 0.05. We tripled the number of individuals to be able to detect associations within three strata ($n = 2{,}400$). Analyses were conducted using R (3.0.2; R Foundation for Statistical Computing) and replicated with Stata 12 (StataCorp). Statistical significance was set at $p < 0.05$.

Results

Children were on average 8.5 y old at baseline, and 50% were girls. The cognitive parameters improved during the 1-y follow-up period (Table 1). On average, working memory increased by 19.0%, superior working memory increased by 15.2%, and inattentiveness decreased by 19.2% (all $p < 0.001$ for linear trend). The magnitude of the 12-mo change was similar in boys and girls, with the exception of superior working memory (numbers), with a lower growth in girls ($p = 0.001$). The cognitive parameters at baseline were negatively associated with maternal education, but not their yearly change (Table 2).

Traffic-related air pollution levels were highly variable between schools (Table 3). EC levels were similar outdoors and indoors, while outdoor levels of NO_2 and UFP were higher than indoor levels. EC showed a high penetration into the classrooms (indoor/outdoor ratio 94.1% [95% CI 85.7%–102.4%]), which was lower for NO_2 (64.5% [95% CI 59.3%–69.7%]) and UFP (70.4% [95% CI 63.5%–77.3%]). Outdoor NO_2 levels at schools were higher than urban background levels. Both during the warm and cold seasons, EC and NO_2 had strong indoor–outdoor correlations, while the correlation was moderate for UFP (Table 4). EC had a strong correlation with NO_2 and with UFP during the warm and cold seasons both outdoors and indoors. EC indoors and UFP outdoors showed the highest correlation between the two seasons. In relation to the covariates, EC and NO_2 were not correlated with the socioeconomic vulnerability index of the school ($r = 0.10$ and 0.00 for EC and −0.08 and −0.15 for NO_2 for outdoors and indoors, respectively, all $p > 0.30$). Correlations between modeled BC and NO_2 at home and measured EC and NO_2 at school were weak ($r = 0.27$, $p < 0.001$, and $r = 0.35$, $p < 0.001$, respectively). Noise was moderately correlated with traffic pollutants ($r = 0.46$, $p = 0.01$, and $r = 0.43$, $p = 0.01$, for indoor EC and NO_2, respectively).

High- and low-exposed schools were comparable in terms of socioeconomic status, although low exposed schools had a higher socioeconomic vulnerability index (i.e., more deprived), were more likely to be public, had higher greenness, and were farther from the busy roads than high-exposed schools (Table 5). Quality of education was identical. However, children attending low-exposed schools had slightly better maternal education; had less behavioral problems, obesity, and foreign origin; had more siblings and residential greenness; and lived farther from the school and commuted less by walking than children from high-polluted schools (Table 5).

Association of High Versus Low Traffic Exposure with Cognitive Development

The difference in 12-mo change in working memory between the low- and high-exposed schools was statistically significant (Table 6). At baseline the difference in working memory

Table 1. Description of the cognitive outcomes in children.

Visit	*n*	Age (Mean)	Working Memory (Two-Back Numbers, *d'*)	Superior Working Memory (Three-Back Numbers, *d'*)	Inattentiveness (HRT-SE, Milliseconds)
1	2,511	8.5 y	221 (131, 363)	112 (59, 188)	267 (202, 336)
2	2,593	8.7 y	222 (131, 392)	123 (59, 190)	248 (184, 318)
3	2,518	9.1 y	236 (131, 392)	129 (59, 190)	243 (181, 314)
4	2,447	9.4 y	263 (153, 392)	129 (64, 212)	224 (163, 291)

Data are median (25th, 75th percentiles).

doi:10.1371/journal.pmed.1001792.t001

Table 2. Cognitive outcomes by maternal education.

Cognitive Outcome	Non-University (n = 1,125)	University (n = 1,590)	p-Value‡
Working memory (two-back numbers, d')			
Baseline	207 (128)	239 (122)	<0.001
12-mo change	30 (161)	29 (153)	0.759
Superior working memory (three-back numbers, d')			
Baseline	108 (100)	127 (100)	<0.001
12-mo change	18 (132)	20 (130)	0.746
Inattentiveness (HRT-SE, milliseconds)			
Baseline	283 (92)	263 (88)	<0.001
12-mo change	–34 (93)	–41 (86)	0.055

Data are mean (standard deviation).
‡Kruskal-Wallis test.

doi:10.1371/journal.pmed.1001792.t002

between low- and high-exposure schools was 5.3 points, while after 1 y this difference had increased to 9.9 points (Table 6), which represents a 4.1% (95% CI 1.5%–6.8%, p = 0.0024) increase in the difference in working memory. Thus, children from high-air-polluted schools had lower improvement in cognitive development compared to children from the paired low-polluted schools (e.g., 7.4%, 95% CI 5.6%–8.8%, versus 11.5%, 95% CI 8.9%–12.5%, 12-mo increase in working memory) (Fig. 2). Similar effects were found for the other cognitive parameters (Fig. 3).

Association of Direct Measurements of Traffic Air Pollution with Cognitive Development

Table 6 gives the adjusted air pollution coefficients at baseline and per 12-mo change for all the cognitive parameters. Children attending schools with higher levels of EC, NO_2, and UFP both in the courtyard and in the classroom had worse cognitive parameters at baseline than children attending schools with lower air pollution. All the coefficients were negative for working memory and positive for inattentiveness, indicating impairment, though the differences were not statistically significant. The growth in cognitive parameters during the 1-y follow-up was also reduced in the schools exposed to higher air pollution levels, which in consequence amplified

Table 3. Description of the air pollutants at the 39 schools.

School Air Pollutant	Minimum	Percentile			Maximum
		25th	50th	75th	
EC outdoor	0.58	1.03	1.32	1.73	3.89
EC indoor	0.44	0.86	1.26	1.78	3.47
NO_2 outdoor	25.9	35.1	48.5	57.4	84.5
NO_2 indoor	11.5	20.5	29.8	38.6	65.6
UFP outdoor	11,939	16,27	22,157	28,257	51,146
UFP indoor	8,034	11,096	14,407	19,968	26,665

Units are micrograms per cubic meter (EC and NO_2) or number per cubic centimeter (UFP). Median NO_2 at the reference urban background station = 41 μg/m^3.

doi:10.1371/journal.pmed.1001792.t003

Table 4. Correlation coefficients (Spearman) between air pollutants by season.

	EC (out)	NO_2 (out)	UFP (out)	EC (in)	NO_2 (in)	UPF (in)
EC (out)	0.58***	0.73***	0.62***	0.82***	0.53**	0.49**
NO_2 (out)	0.63***	0.49**	0.51**	0.61***	0.71***	0.34
UPF (out)	0.61***	0.61***	0.72***	0.49**	0.30	0.57***
EC (in)	0.86***	0.69***	0.63***	0.73***	0.66***	0.61***
NO_2 (in)	0.45**	0.70***	0.43*	0.58***	0.64***	0.39*
UFP (in)	0.41*	0.42*	0.65***	0.62***	0.38*	0.40*

Below diagonal, cold season (November–March); above diagonal, warm season (April–October). Correlations between the two seasons in the diagonal.
*$p < 0.05$
**$p < 0.01$
***$p < 0.001$.
out, outdoors (courtyard); in, indoors (classroom).

doi:10.1371/journal.pmed.1001792.t004

the differences between schools at the end of follow-up. The detrimental association of air pollution with change in the cognitive parameters was observed for all the outcomes and pollutants, being statistically significant for almost all of them. Thus, for example, after 1 y of follow-up, the difference in working memory for a change from the first to the fourth quartile of indoor EC had increased by 6.2 (95% CI 2.0–11.0) points ($p = 0.004$) (13.0% [95% CI 4.2%–23.1%] of the total growth). When the stimulus was words instead of numbers, the results were very similar for superior working memory (Table 7). Fig. 4 shows the change in working memory in 1 y as a function of both outdoor and indoor pollutant levels. The points in the figure represent the crude estimates of change in cognitive parameters for each school along with the school air pollution levels, while the line represents the regression line obtained from the final adjusted model. Fig. 4 illustrates the negative relationship between change in cognitive function and air pollution levels, and depicts a good fit between the crude values and the adjusted slope. Similar findings were seen for the other cognitive parameters (Figs. 5 and 6).

Sensitivity Analyses

The crude and the adjusted models with high- versus low-air-pollution schools and with the direct measures of air pollutants gave similar results (Fig. 3; S1 Table). Further adjustment for the individual socioeconomic factors included in Table 5, ADHD or behavioral symptoms, residential greenness, and school noise and greenness did not materially change associations between high/low air pollution; EC, NO_2, and UFP; and 12-mo change in cognitive parameters. Similarly, results remained unchanged after adjusting for high-/low-air-pollution area, commuting, smoking at home (S2 Table), educational quality, and participation rate per school.

In stratified analysis, associations of cognitive parameters with EC (Table 8), NO_2, and UFP were similar in high-air-pollution schools and low-air-pollution schools, as well as according to neighborhood socioeconomic status and obesity. In contrast, detrimental associations were stronger in general in boys than in girls, in children from more highly educated mothers, in children from private schools, and in children with ADHD symptoms, though differences were not significant (p for interaction > 0.1 in the mixed effects linear models), and the detrimental associations occurred in all the groups. Given that development was significantly lower in grade 4 for all tasks, we repeated the analyses stratifying by grade, and the results were homogeneous. Moreover, in order to control for the "summer learning loss" phenomenon occurring between the two academic years, we excluded tests done in the second academic year that did

Table 5. Population and school characteristics by school traffic (from original design).

Characteristic	Low Traffic	High Traffic	*p*-Value‡
Schools			
Number	20	19	
School socioeconomic vulnerability index	0.52 (0.24)	0.41 (0.16)	0.055
School greenness (NDVI)	0.31 (0.10)	0.15 (0.03)	<0.001
Type of school, public	55%	42%	0.421
Educational quality (PISA 2012)	3.9 (1.3)	3.9 (1.8)	0.790
Noise level in classroom (decibels)	37.2 (4.9)	40.1 (5.0)	0.068
Distance to busy road (meters)	369 (357)	118 (178)	<0.001
EC outdoor ($\mu g/m^3$)	1.13 (0.39)	1.82 (0.70)	<0.001
NO_2 outdoor ($\mu g/m^3$)	40.5 (9.6)	56.1 (11.5)	<0.001
UFP outdoor (number/cm^3)	18,043 (5,702)	28,745 (8,326)	0.001
Children			
Number	1,355	1,360	
Girls	49%	51%	0.318
Foreign origin (non-Spanish)	11%	19%	<0.001
Maternal education, university	62%	55%	<0.001
Paternal education, university	58%	48%	<0.001
Maternal occupation, unemployed	17%	19%	0.036
Paternal occupation, unemployed	8%	12%	<0.001
Marital status, married	86%	84%	0.053
Home socioeconomic vulnerability index	0.43 (0.22)	0.47 (0.19)	<0.001
Home greenness (NDVI)	0.022 (0.09)	0.017 (0.005)	<0.001
Commuting to school, walking	33%	73%	<0.001
Distance from home to school (meters)	2,430 (2,359)	1,028 (1,577)	<0.001
Behavioral problems (SDQ)	7.9 (5.0)	8.9 (5.4)	<0.001
Overweight/obese	25%	30%	0.002
Computer games weekend, ≥1 h	69%	72%	0.081
Siblings, yes	83%	75%	<0.001
Adopted child	4%	4%	0.793
Secondhand smoke at home	12%	14%	0.069
Smoking during pregnancy	10%	10%	0.785
Gestational age < 37 wk	8%	7%	0.497
Birth weight < 2,500 g	9%	10%	0.994
Breastfeeding, no	18%	18%	0.272

Data are number, percent, or mean (standard deviation).

‡Kruskal-Wallis and Chi-square tests.

PISA, Programme for International Student Assessment.

doi:10.1371/journal.pmed.1001792.t005

not result in a notable change in our observed associations. Furthermore, we excluded the first exam, to prevent a potential practice effect, and the association, if anything, became stronger for working memory and superior working memory (S3 Table). Finally, sequential exclusion of school pairs one by one from the models did not change the results, suggesting that exceptional influential cases were not affecting the results.

Table 6. Difference in cognitive development, at baseline and 12-mo change, by school air pollution exposure (high/low group or interquartile range increase) in 2,715 children and 10,112 tests from 39 schools.

Cognitive Outcome	High/Low Traffic	Outdoor (Courtyard)			Indoor (Classroom)		
		EC	NO_2	UFP	EC	NO_2	UFP
Units per interquartile range	—	0.7 μ g/m^3	23.3 μg/m^3	6,110 counts	0.92 μg/m^3	18.1 μg/m^3	8,872 counts
Working memory (two-back numbers, *d'*)							
Baseline	−5.3 (−16, 5.1)	−5.8 (−12, 0.56)	−7.5 (−16, 0.99)	−6.4 (−14, 1.5)	−3.0 (−11, 4.8)	−6.1 (−14, 1.9)	−1.3 (−13, 9.9)
12-mo change	−9.9 (−16, −3.5)*	−4.1 (−8.0, −0.2)*	−6.6 (−12, −1.2)*	−4.9 (−10, 0.22)	−6.2 (−11, −2.0)*	−5.6 (−11, −0.44)*	−7.9 (−15, −1.3)*
Superior working memory (three-back numbers, *d'*)							
Baseline	−1.4 (−10, 7.1)	0.25 (−5.2, 5.7)	1.5 (−5.8, 8.8)	−0.95 (−7.4, 5.6)	1.4 (−5.0, 7.9)	1.3 (−5.4, 8.0)	−0.078 (−9.1, 8.9)
12-mo change	−5.8 (−11, −0.74)*	−4.4 (−7.6, −1.3)*	−6.7 (−11, −2.3)*	−5 (−9.1, −0.96)*	−5.8 (−9.2, −2.4)*	−5.1 (−9.2, −0.91)*	−6.0 (−11, −0.75)*
Inattentiveness (HRT-SE, milliseconds)							
Baseline	5.2 (−6.2, 17)	1 (−6.3, 8.4)	4.8 (−5.0, 14)	4.5 (−4.0, 13)	6.8 (−1.7, 15)	7.0 (−1.8, 16)	6.2 (−5.8, 18)
12-mo change	5.2 (0.68, 9.7)*	3.8 (1.0, 6.6)*	3.8 (−0.10, 7.6)	3.9 (0.31, 7.6)*	3.8 (0.79, 6.8)*	2.6 (−1.0, 6.3)	4.6 (−0.13, 9.2)

Difference (95% CI) in the 12-mo change adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects.

*$p < 0.05$.

doi:10.1371/journal.pmed.1001792.t006

Discussion

This large study with repeated and objective measures demonstrated that cognitive development is reduced in children exposed to higher levels of traffic-related air pollutants at school. This association was consistent for working memory, superior working memory, and inattentiveness, and robust to several sensitivity analyses. The association was observed both when the exposure was treated as high/low traffic-related air pollution and when using specific pollutants including outdoor and indoor EC, NO_2, and UFP, which are largely traffic-related [21,22]. Changes in the developmental trajectory could resemble those suggested for the adverse impact of urban air pollution on lung function development [29]. Mechanisms of air-pollution-induced neurotoxicity have been explored [30]. The findings provide strong support for air pollution being a developmental neurotoxicant and point towards the primary school age as a particularly vulnerable time window for executive function development.

A strength of this study is the longitudinal ascertainment of executive function trajectories that specifically develop during school age and the direct measures of air pollution. A concern, however, is potential residual confounding by socio-demographic characteristics, although in European cities, the relationship between proximity to traffic and economically disadvantaged areas is not always evident [31]. In the city of Barcelona, the highest air pollution was observed in the "Eixample," a wealthy central area of the city where most of our schools with high traffic were selected [23]. We paired by design high- and low-traffic schools by socioeconomic characteristics and type of school, and although there was an inverse relation between school pollution and socioeconomic vulnerability index, such differences between schools after matching became small. In addition to the association of cognitive parameters observed with high- compared to low-exposed schools, we also observed a consistent association of cognitive parameters with specific pollutants whose relation with socio-demographics was weak and in some cases nonexistent. Furthermore, cognitive development was unrelated to social determinants in our study, in contrast to cognitive function at baseline. Besides, the associations remained in

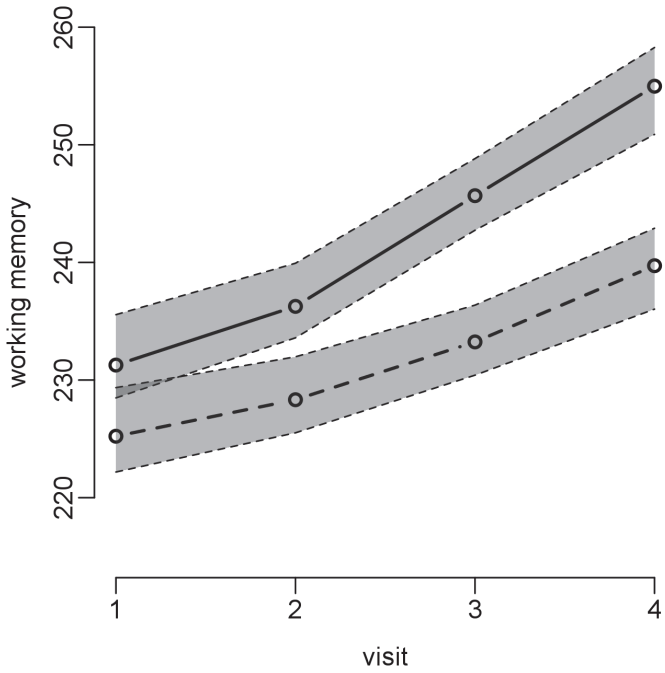

Fig 2. Working memory development by high- or low-traffic-air-pollution school. Dashed line = high traffic air pollution; continuous line = low traffic air pollution; gray shading indicates 95% CIs. Adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects in 2,715 children and 10,112 tests from 39 schools.

doi:10.1371/journal.pmed.1001792.g002

the stratified analyses (e.g., for type of school or high-/low-polluted area) and after additional adjustment (e.g., for commuting, educational quality, or smoking at home), contradicting a potential residual confounding explanation.

Other potential limitations are the potential misclassification error of the UFP exposures. Seasonalized measures of UFP showed the lowest correlation among the pollutants between the first and the second campaign and weaker associations with the cognitive parameters (e.g., −4.0 [95% CI −8.6 to 0.49] for indoor UFP and working memory) than non-seasonalized UFP, which is probably because of its large geographical and temporal instability due to constant and rapid secondary formation [22]. In contrast, EC and NO_2 showed very similar associations with cognitive parameters using both seasonalized and non-seasonalized measures. Another potential limitation is non-response. A total of 182 out of the initial 2,897 children (6%) were excluded because of incomplete data on individual variables. When these children were included in the analysis in models that did not require the complete dataset (i.e., a model not adjusted for maternal education), results were identical. Another level of non-response refers to children (41%) from families that did not want to be part of the study, although they were invited. This non-response affects representativeness rather than internal validity, given that the participation rate per school was unrelated to the school social gradient and that adjustment for participation rate did not change the results. Based on the results from one school, participants had less neuropsychological problems than non-participants, which likely made them less susceptible to air pollution effects. Therefore, any effect observed in the present study

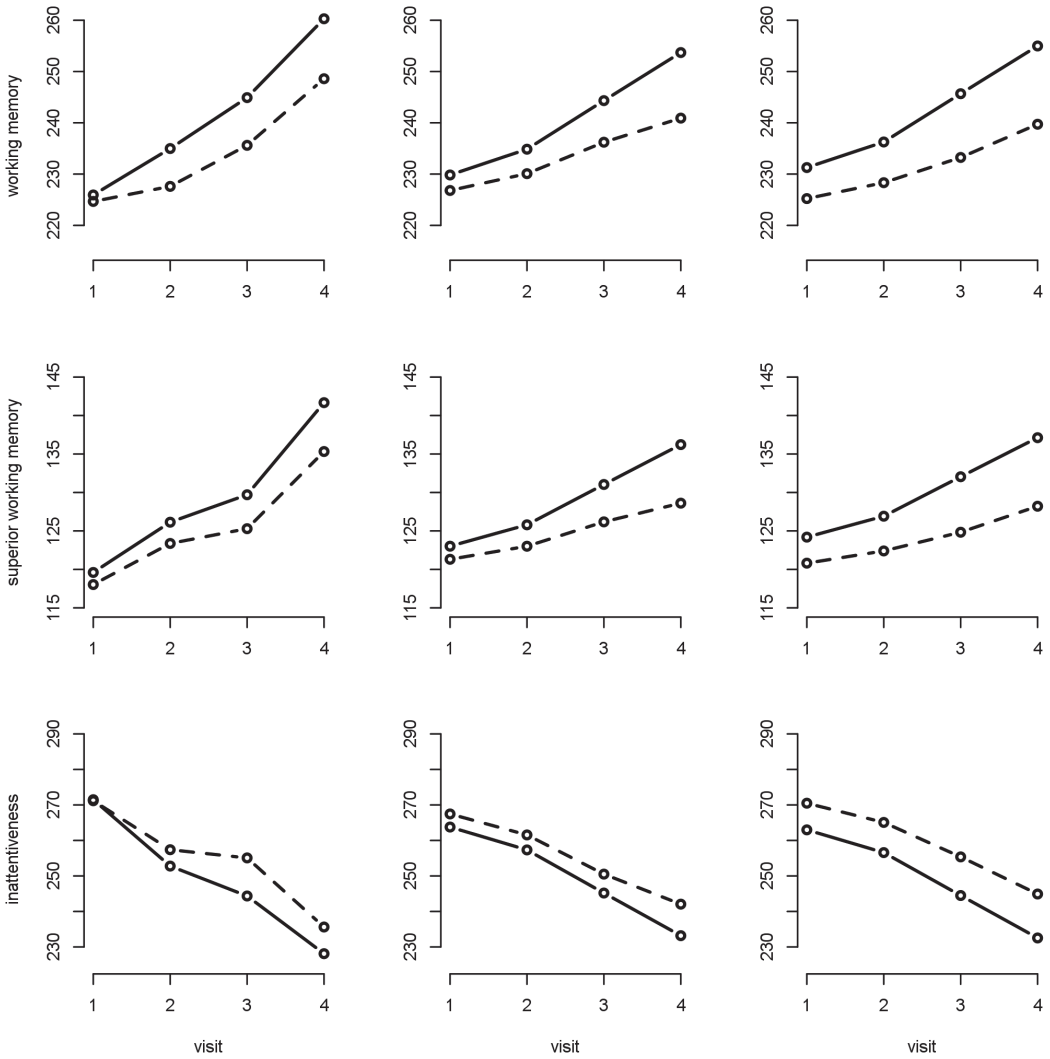

Fig 3. Crude and adjusted cognitive development by high- or low-air-pollution school. Dashed line = high air pollution; continuous line = low air pollution. The first column depicts the crude values, the second the crude trajectories from a model that included individual and school as random effects, and the third a model adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects in 2,715 children and 10,112 tests from 39 schools.

doi:10.1371/journal.pmed.1001792.g003

would likely be a conservative estimate for extrapolation to the entire population. A third limitation relates to the lack of measurements in preceding periods. However, all children had been in their school for more than 6 mo before the beginning of the study, and when we limited the study to children with more than 2 y in the school (94% of the children), associations remained the same. We interpreted these associations as chronic effects (i.e., due to exposures longer than 6 mo) since it is unlikely that the geographical pattern of air pollution occurring during

Table 7. Difference in cognitive development for tests using words, at baseline and 12-mo change, by school air pollution exposure (high/low group or interquartile range increase) in 2,715 children and 10,112 tests from 39 schools.

Cognitive Outcome	High/Low Traffic	Outdoor (Courtyard)			Indoor (Classroom)		
		EC	NO_2	UFP	EC	NO_2	UFP
Working memory (two-back words, *d'*)							
Baseline	−8.4 (−19, 1.9)	−3.7 (−10, 2.9)	−3.3 (−12, 5.6)	−5.0 (−13, 3.0)	−4.1 (−12, 3.5)	−4.3 (−12, 3.8)	−4.3 (−15, 6.4)
12-mo change	−4.7 (−11, 1.7)	−1.7 (−5.6, 2.3)	−3.4 (−8.9, 2.1)	−3.1 (−8.2, 2.0)	−2.3 (−6.6, 2.0)	0.60 (−4.6, 5.8)	−5.4 (−12, 1.2)
Superior working memory (three-back words, *d'*)							
Baseline	−1.8 (−8.5, 4.9)	0.25 (−4.0, 4.5)	0.96 (−4.8, 6.7)	−0.67 (−5.9, 4.6)	0.88 (−4.0, 5.7)	0.096 (−5.2, 5.4)	0.40 (−6.7, 7.5)
12-mo Change	−5.9 (−11, −0.89)*	−4.9 (−8.0, −1.8)*	−6.8 (−11, −2.5)*	−5.7 (−9.7, −1.7)*	−5.4 (−8.7, −2.0)*	−3.9 (−8.0, 0.14)	−3.9 (−9.1, 1.3)

Difference (95% CI) in the yearly change adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects.

*$p < 0.05$

doi:10.1371/journal.pmed.1001792.t007

the study period had changed in the last 2 y. Finally, indoor assessment was limited to a single classroom. This is not a problem for the indoor assessment of pollutants such as EC, given the high correlation between outdoor and indoor levels and similar coefficients for the association with cognition between outdoor and indoor exposures. However, it could be a problem for school noise since the correlation between outdoor and indoor noise was strongly dependent on the street orientation of the classroom (ranging from 0.07 for classrooms facing away from the street to 0.70 for classrooms facing the street). However, residual confounding by noise was unlikely given the weak correlation between the pollutants and noise measured in the same classrooms, and the robustness of the coefficients for the different pollutants after adjusting for noise and for the interaction between noise and age.

This study addresses the role of traffic air pollution in schools on cognitive development. Previous studies on the effects of polluted air at schools were a study in two schools in Quanzhou (China) on attention disorders [10], two studies on aircraft noise that secondarily assessed the association between NO_2 and cognitive function [32,33], and an ecological study in Michigan (US) on industrial pollution and school failure [34]. Other studies in children have evaluated the effect of maternal personal air pollution exposure or maternal/child exposure at home [35]. We found here an association between traffic-related air pollution exposure at school and cognitive development during primary school age, independent of residential air pollution and beyond the effects related to home exposures in early life found by previous studies. Total cumulative exposure in school, home, and commuting and the different time windows of exposure are not addressed here, but the continuous monitoring of BC and physical activity with personal samplers in 54 of our children showed that exposure at school was significantly higher than at home and did not change by commuting mode. This higher exposure level at school could be attributed to peaks of pollution occurring during school time, and higher inhaled dose during school time due to exercise and physical activity at schools. Besides, the fact that children at schools in the most polluted area traveled a shorter distance from home suggests a shorter commute, which could explain the lack of confounding after adjusting for commuting distance and mode. We could not disentangle the time frame of the exposures occurring under the long-term school exposure measured here. However, in the case of inattentiveness, in contrast to what was seen for working memory, the association at baseline was larger than at follow-up. Given that inattentiveness develops earlier than working memory

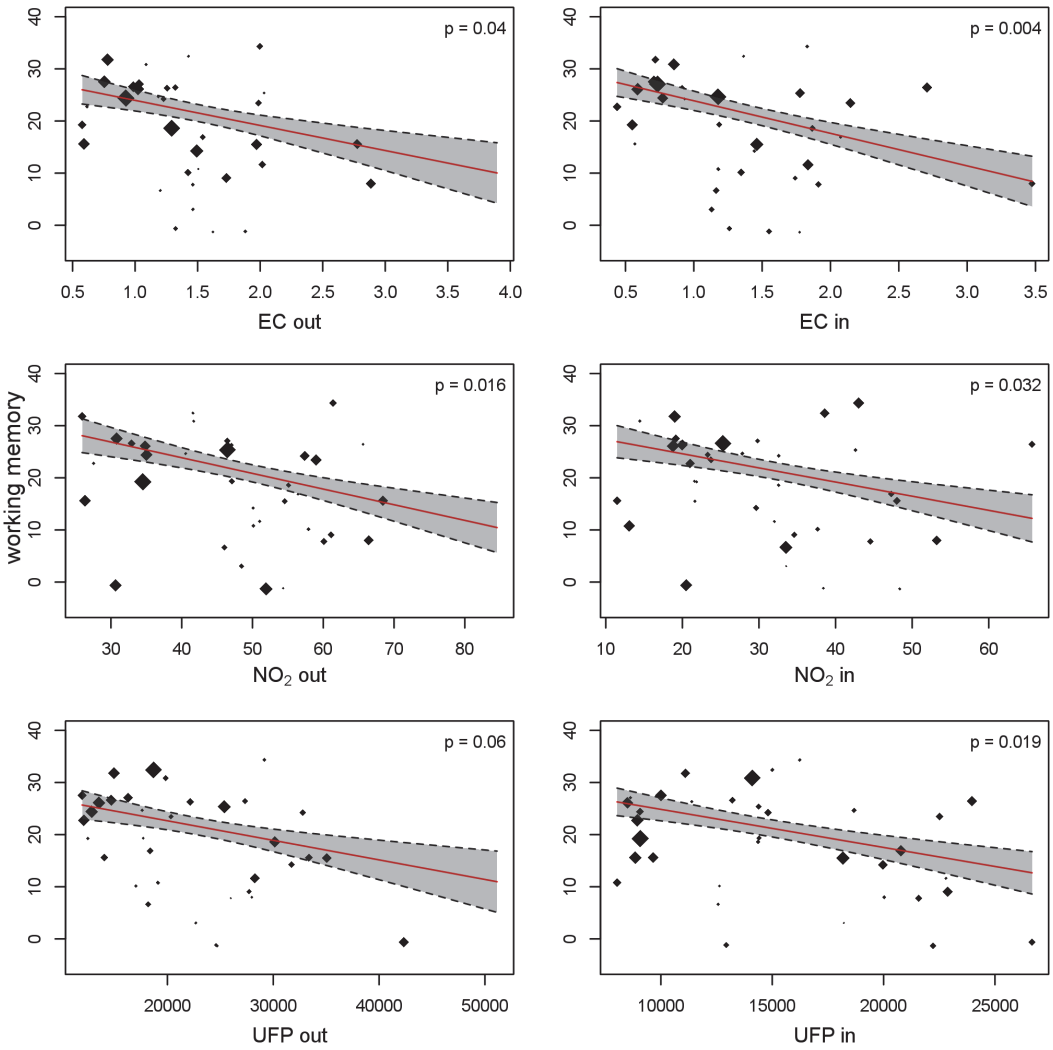

Fig 4. Working memory development and long-term exposure to traffic-related air pollutants. Each dot depicts a school, with size proportional to the number of children. The cognitive development per school was estimated in a model with school and individual as random effects. The slope of the red line depicts the change in cognitive development as a function of the air pollutants, adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects in 2,715 children and 10,112 tests from 39 schools. Gray shading indicates 95% CIs. out, outdoors (courtyard); in, indoors (classroom).

doi:10.1371/journal.pmed.1001792.g004

[12], this finding could suggest that the adverse effect of air pollution could have preceded the study period, and that the lower improvement in scores may be associated with previous exposures, too.

Among the individual traffic-related pollutants, we found an adverse association between EC and child cognitive development. EC comes almost exclusively from diesel vehicles in Barcelona, with an ambient air mode of around 30–40 nm, in the UFP range [22]. EC and traffic-

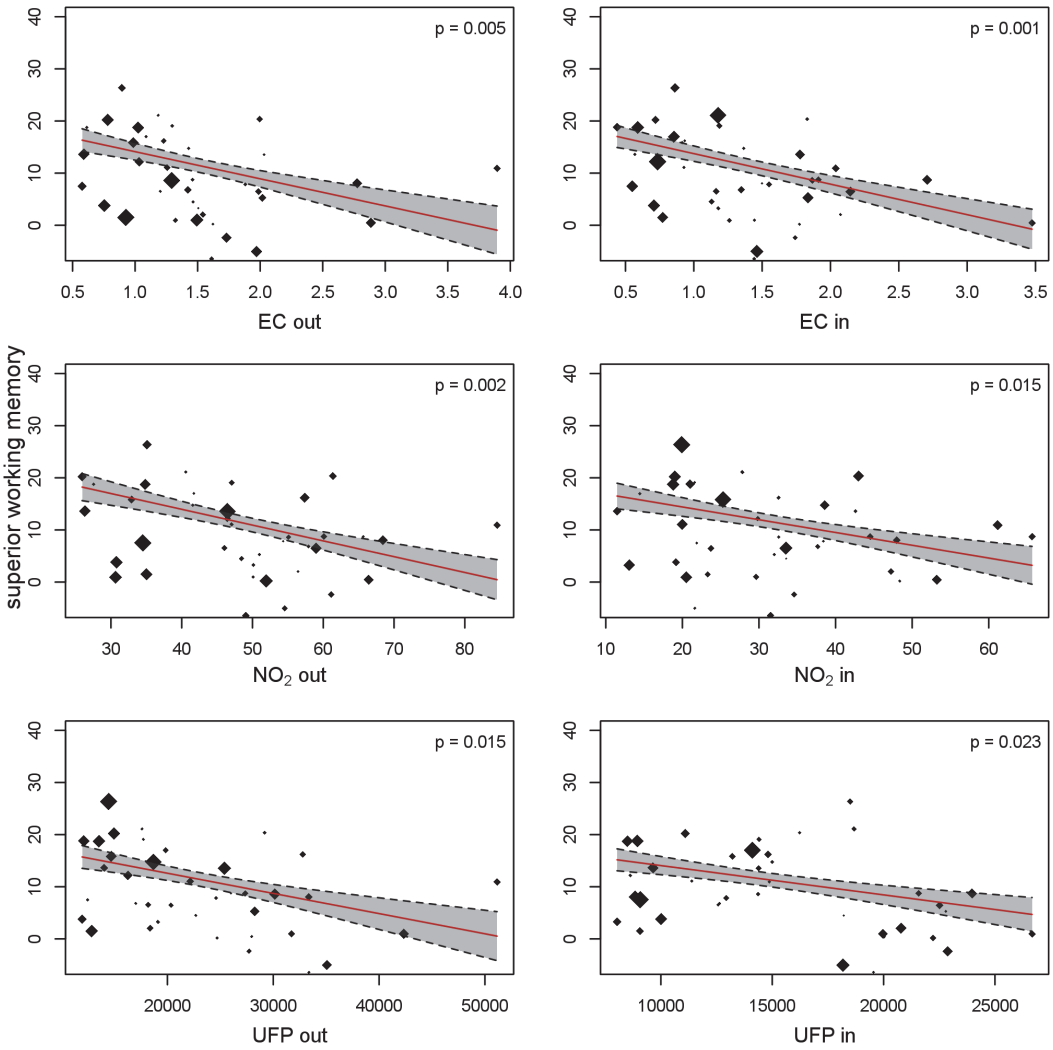

Fig 5. Superior working memory and long-term exposure to traffic-related air pollutants. Each dot depicts a school, with size proportional to the number of children. The cognitive development per school was estimated in a model with school and individual as random effects. The slope of the red line depicts the change in cognitive development as a function of the air pollutants, adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects in 2,715 children and 10,112 tests from 39 schools. Gray shading indicates 95% CI. out, outdoors (courtyard); in, indoors (classroom).

doi:10.1371/journal.pmed.1001792.g005

derived metals were an important fraction of indoor and outdoor quasi-ultrafine particles ($PM_{0.25}$) in our study schools [36]. We observed a high penetration of EC into the classrooms (indoor/outdoor ratio 94%) and similar associations of indoor and outdoor EC with cognitive development. Although the indoor/outdoor ratio was weaker (70%) for UFP, we also found cognitive associations with UFP. These findings remained after adjustment for traffic noise at school, pointing towards UFP as a neurotoxic traffic component, which is coherent with the

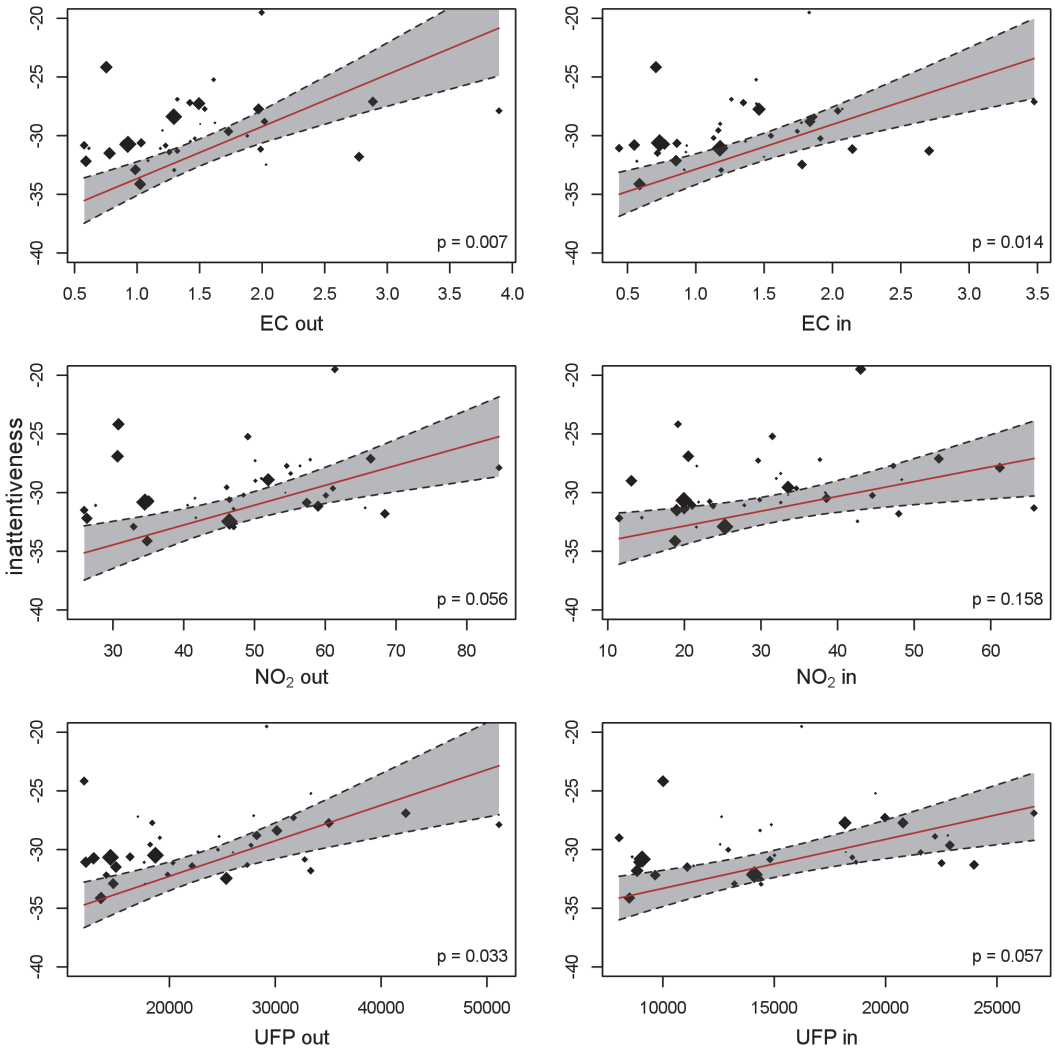

Fig 6. Inattentiveness development and long-term exposure to traffic-related air pollutants. Each dot depicts a school, with size proportional to the number of children. The cognitive development per school was estimated in a model with school and individual as random effects. The slope of the red line depicts the change in cognitive development as a function of the air pollutants, adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects in 2,715 children and 10,112 tests from 39 schools. Gray shading indicates 95% CI. out, outdoors (courtyard); in, indoors (classroom).

doi:10.1371/journal.pmed.1001792.g006

numerous and consistent findings in animal studies that UFP may cause disruption of the blood–brain barrier, microglial activation, and brain inflammation [14].

Evidence points towards chronic microglial stimulation and altered innate immune response and inflammation as the key neurotoxic mechanisms of UFP [14,29,37]. UFP has been shown to cause microglial inflammation following either brain UFP deposition or systemic inflammation originating in UFP-exposed organs such as the lungs [36]. Microglial stimulation

Table 8. Stratified analyses of adjusted 12-mo change in cognitive development by school air pollution exposure (high/low group or interquartile range increase) in 2,715 children and 10,112 tests from 39 schools.

Cognitive Outcome	By Sex		By Maternal Education		By ADHD		By High/Low Air Pollution		By Type of School	
	Boys (n = 1,357)	Girls (n = 1,358)	High (n = 1,590)	Low–Middle (n = 1,125)	No (n = 2,409)	Yes (n = 275)	High (n = 1,358)	Low (n = 1,357)	Public (n = 931)	Private (n = 1,784)
Working memory (two-back numbers, *d'*)										
High/low	−13 (−23, −4.2)*	−6.1 (−15, 2.6)	−15 (−23, −6.4)*	−3.2 (−13, 6.7)	−7.7 (−14, −0.97)*	−26 (−45, −6.7)*	–	–	−0.15 (−12, 11)	−14 (−22, −6.4)*
EC outdoor	−6.4 (−12, −0.75)*	−1.3 (−6.7, 4.0)	−10 (−15, −5.1)*	4 (−2.2, 10)	−1.9 (−6.0, 2.3)	−17 (−29, −5.6)*	1.2 (−4.6, 6.9)	−6.9 (−16, 2.4)	3.9 (−3.0, 11)	−8.0 (−13, −3.1)*
EC indoor	−8.9 (−15, −2.8)*	−3.2 (−9.1, 2.8)	−10 (−16, −4.7)*	−0.64 (−7.5, 6.2)	−3.5 (−8.0, 1.1)	−22 (−35, −8.5)*	−2.7 (−8.8, 3.5)	−6.6 (−18, 5.0)	−0.53 (−11, 10)	−7.1 (−12, −2.3)*
Superior working memory (three-back numbers, *d'*)										
High/low	−10 (−18, −3.0)*	−1.9 (−8.8, 5.0)	−7.5 (−14, −0.74)*	−3.7 (−11, 4.0)	−5.2 (−11, 0.14)	−12 (−26, 3.0)	–	–	−2.1 (−11, 7.1)	−7.3 (−13, −1.2)*
EC outdoor	−9.6 (−14, −5.1)*	1.2 (−3.1, 5.5)	−6.7 (−11, −2.6)*	−1.2 (−6.0, 3.6)	−3.3 (−6.7, 0.03)	−11 (−19, −1.8)*	−3.1 (−7.8, 1.5)	−4.8 (−12, 2.5)	−1.8 (−7.3, 3.7)	−5.5 (−9.4, −1.6)*
EC indoor	−10 (−15, −5.4)*	−0.85 (−5.6, 3.9)	−8.9 (−13, −4.5)*	−1.4 (−6.7, 3.9)	−4.7 (−8.4, −1.1)*	−11 (−20, −0.95)*	−5.7 (−11, −0.71)*	−4.2 (−13, 4.9)	−4.9 (−13, 3.4)	−5.7 (−9.6, −1.9)*
Inattentiveness (HRT-SE, milliseconds)										
High/low	8.1 (1.8, 15)*	1.4 (−4.9, 7.8)	9.0 (3.1, 15)*	−0.93 (−8.2, 6.3)	5.5 (0.69, 10)*	3.6 (−11, 18)	–	–	1.1 (−7.1, 9.2)	7.9 (2.4, 13)*
EC outdoor	5.8 (1.9, 9.6)*	1.8 (−2.2, 5.7)	5.2 (1.7, 8.7)*	1.4 (−3.0, 5.9)	2.3 (−0.63, 5.2)	13 (4.9, 22)*	4.7 (0.72, 8.8)*	−2 (−8.6, 4.5)	3.6 (−1.1, 8.3)	4.5 (1.0, 8.0)*
EC indoor	5.2 (1.0, 9.4)*	2.0 (−2.3, 6.4)	4.6 (0.84, 8.4)*	1.9 (−3.1, 6.8)	1.9 (−1.3, 5.2)	16 (7.0, 26)*	3.9 (−0.47, 8.2)	−2 (−10, 6.2)	5.3 (−2.0, 13)	4.2 (0.74, 7.6)*

Difference (95% CI) in the 12-mo change, adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects.

*$p < 0.05$.

doi:10.1371/journal.pmed.1001792.t008

affects neurons, and UFP has been shown to decrease neuronal glutamatergic function and disrupt synapses [38]. Similarly, airborne metals have been shown to alter dopamine function [39]. The underlying brain mechanisms are beyond the present study, but the observation of associations with executive functions, the lack of confounding by ADHD or behavior, and the association among children without ADHD suggests a general brain dysfunction.

Boys appeared more susceptible to air pollution, although both boys and girls showed an adverse association of school air pollution with cognitive development. Although results could be due to chance, in animals, males were more susceptible to airborne metals than females, which may be because of sex-specific altered dopamine function [39]. The possible higher vulnerability of children with ADHD could also indicate abnormalities related to dopamine [40]. Stratification by maternal education or type of school showed a larger association among students with the most educated mothers and those from private schools. This resembles what has been observed with other hazards for neurodevelopment such as genetic effects [41], presumably because fewer adverse factors play a role among students with educated mothers or in private schools, thus causing less interference with the factors under study.

The observed association between air pollution and cognitive development was strong. For example, an increase from the first to the fourth quartile in indoor EC resulted in a 13.0% reduction in the growth of working memory. In contrast, the association at baseline was smaller (1.9%). Part of this larger association during primary school may be a matter of bigger magnitude of exposure to traffic pollution in schools, but it could indicate that some executive functions are particularly vulnerable during primary school age, as has also been seen for lead [42]. The long-term effect probably occurs over the period of maximum development of these executive functions, resulting in a notable cumulative effect by the end of this period in preadolescence. The observed association was consistent across cognitive measurements, though it was more evident for superior working memory, which is a good predictor of learning achievement [19]. Impairment of high cognitive functions has severe consequences for school achievement [11]. Thus, reduced cognitive development in children attending the most polluted schools might result in a disadvantage in mental capital, which may have a long-lasting life course effect.

Overall, we have shown that children attending schools with higher levels of exposure to traffic-related air pollutants had a smaller growth in cognitive development over time, suggesting that traffic-related air pollution in schools negatively affects cognitive development. This may have consequences for learning, school achievement, and behavior. With regard to air pollution regulation, the present study shows that the developing brain may be vulnerable to certain traffic-related air pollutants.

Supporting Information

S1 Table. Crude difference (and 95% CI) in cognitive development at baseline and 12-mo change by school air pollution exposure (high versus low or interquartile range increase) in 2,715 children and 10,112 tests from 39 schools.
(DOCX)

S2 Table. Difference (and 95% CI) in cognitive development at baseline and 12-mo change by school air pollution exposure (high versus low or interquartile range increase) in 2,715 children and 10,112 tests from 39 schools, after further adjustment for high/low area, commuting, and smoking at home.
(DOCX)

S3 Table. Difference (and 95% CI) in cognitive development (12-mo change) by school air pollution exposure (high/low group or interquartile range increase) in 2,715 children and 10,112 tests, after excluding some child-visits.
(DOCX)

S1 Text. STROBE checklist.
(DOCX)

Acknowledgments

We acknowledge all the families participating in the study for their altruism and particularly to the schools Antoni Brusi, Baloo, Betània–Patmos, Centre d'Estudis Montseny, Col·legi Shalom, Costa i Llobera, Escola El Sagrer, Els Llorers, Escola Pia de Sarrià, Escola Pia Balmes, Escola Concertada Ramon Llull, Escola Lourdes, Escola Tècnica Professional del Clot, Ferran i Clua, Francesc Macià, Frederic Mistral, Infant Jesús, Joan Maragall, Jovellanos, La Llacuna del Poblenou, Lloret, Menéndez Pidal, Nuestra Señora del Rosario, Miralletes, Ramon Llull, Rius i Taulet, Pau Vila, Pere Vila, Pi d'en Xandri, Projecte, Prosperitat, Sant Ramon Nonat–Sagrat Cor, Santa Anna, Sant Gregori, Sagrat Cor Diputació, Tres Pins, Tomàs Moro, Torrent d'en Melis, and Virolai. We also acknowledge the ESCAPE project for the design and supervision modeling of air pollution. Furthermore, Xavier Mayoral did the technical development of the *n*-back test; Cecilia Persavento, Judit Gonzalez, Laura Bouso, and Pere Figueras contributed to the field work; and Michelle Turner contributed to the copyediting.

Author Contributions

Conceived and designed the experiments: JS XQ. Performed the experiments: MAP JF MLV ESG MF IR MV TM AA MN NSG. Analyzed the data: ME RGE XB MC MAP MF. Contributed reagents/materials/analysis tools: MN ME XB. Wrote the paper: JS ME MAP JF IR MLV ESG MF RGE XB MV MC TM AA NSG MN XQ. Agree with manuscript results and conclusions: JS ME MAP JF IR MLV ESG MF RGE XB MV MC TM AA NSG MN XQ. All authors have read, and confirm that they meet, ICMJE criteria for authorship.

References

1. Grandjean Ph, Landrigan Ph J (2014) Neurobehavioural effects of developmental toxicity. Lancet Neurol 13: 330–338. doi: 10.1016/S1474-4422(13)70278-3 PMID: 24556010
2. Gerlofs-Nijland ME, van Berlo D, Cassee FR, Schins RP, Wang K, et al. (2010) Effect of prolonged exposure to diesel engine exhaust on proinflammatory markers in different regions of the rat brain. Part Fibre Toxicol 7: 12. doi: 10.1186/1743-8977-7-12 PMID: 20478040
3. Bos I, De Boever P, Int Panis L, Sarre S, Meeusen R (2012) Negative effects of ultrafine particle exposure during forced exercise on the expression of Brain-Derived Neurotrophic Factor in the hippocampus of rats. Neuroscience 223: 131–139. doi: 10.1016/j.neuroscience.2012.07.057 PMID: 22867973
4. Yokota S, Takashima H, Ohta R, Saito Y, Miyahara T, et al. (2011) Nasal instillation of nanoparticle-rich diesel exhaust particles slightly affects emotional behavior and learning capability in rats. J Toxicol Sci 36: 267–276. PMID: 21628955
5. Fonken LK, Xu X, Weil ZM, Chen G, Sun Q, et al. (2011) Air pollution impairs cognition, provokes depressive-like behaviors and alters hippocampal cytokine expression and morphology. Mol Psychiatry 16: 987–995. doi: 10.1038/mp.2011.76 PMID: 21727897
6. Suglia SF, Gryparis A, Wright RO, Schwartz J, Wright RJ (2008) Association of black carbon with cognition among children in a prospective birth cohort study. Am J Epidemiol 167: 280–286. PMID: 18006900
7. Perera FP, Li Z, Whyatt R, Hoepner L, Wang S, et al. (2009) Prenatal airborne polycyclic aromatic hydrocarbon exposure and child IQ at age 5 years. Pediatrics 124: e192–e202. doi: 10.1542/peds.2009-0334 PMID: 19706576

8. Guxens M, Aguilera I, Ballester F, Estarlich M, Fernández-Somoano A, et al. (2012) Prenatal exposure to residential air pollution and infant mental development: modulation by antioxidants and detoxification factors. Environ Health Perspect 120: 144–149. doi: 10.1289/ehp.1103469 PMID: 21868304
9. McConnell R, Islam T, Shankardass K, Jerrett M, Lurmann F, et al. (2010) Childhood incident asthma and traffic-related air pollution at home and school. Environ Health Perspect 118: 1021–1026. doi: 10.1289/ehp.0901232 PMID: 20371422
10. Wang S, Zhang J, Zeng X, Zeng Y, Wang S, et al. (2009) Association of traffic-related air pollution with children's neurobehavioral functions in Quanzhou, China. Environ Health Perspect 117: 1612–1618. doi: 10.1289/ehp.0800023 PMID: 20019914
11. Alloway TP, Alloway RG (2010) Investigating the predictive roles of working memory and IQ in academic attainment. J Exp Child Psychology 106: 20–29.
12. Anderson P (2002) Assessment and development of executive function (EF) during childhood. Child Neuropsychol 8: 71–82. PMID: 12638061
13. Thomason ME, Race E, Burrows B, Whitfield-Gabrieli S, Glover H, et al. (2009) Development of spatial and verbal working memory capacity in the human brain. J Cogn Neurosci 21: 316–332. doi: 10.1162/jocn.2008.21028 PMID: 18510448
14. Block ML, Calderón-Garcidueñas L (2009) Air pollution: mechanisms of neuroinflammation and CNS disease. Trends Neurosci 32: 506–516. doi: 10.1016/j.tins.2009.05.009 PMID: 19716187
15. Wang M, Beelen R, Basagana X, Becker T, Cesaroni G, et al. (2013) Evaluation of land use regression models for NO2 and particulate matter in 20 European study areas: the ESCAPE project. Environ Sci Technol 47: 4357–4364. doi: 10.1021/es305129t PMID: 23534892
16. Rueda MR, Rothbart MK, McCandliss BD, Saccomanno L, Posner MI (2005) Training, maturation, and genetic influences on the development of executive attention. Proc Natl Acad Sci U S A 102: 14931–14936. PMID: 16192352
17. Rueda MR, Fan J, McCandliss BD, Halparin J D, Gruber DB, et al. (2004) Development of attentional networks in childhood. Neuropsychology 42: 1029–1040.
18. Forns J, Esnaola M, López-Vicente M, Suades-González E, Alvarez-Pedrerol M, et al. (2014) The n-back test and the attentional network task as measures of child neuropsychological development in epidemiological studies. Neuropsychology 28: 519–529. doi: 10.1037/neu0000085 PMID: 24819069
19. Shelton JT, Elliott EM, Matthews RA, Hill BD, Gouvier WD (2010) The relationships of working memory, secondary memory, and general fluid intelligence: working memory is special. J Exp Psychol Learn Mem Cogn 36: 813–820. doi: 10.1037/a0019046 PMID: 20438278
20. Conners CK, Multi-Health Systems (2000) Conners' Continuous Performance Test II: computer program for Windows technical guide and software manual. North Tonwanda (New York): Multi-Health Systems.
21. Amato F, Rivas I, Viana M, Moreno T, Bouso L, et al. (2014) Sources of indoor and outdoor PM2.5 concentrations in primary schools. Science Total Environ 490: 757–765. doi: 10.1016/j.scitotenv.2014.05.051 PMID: 24907610
22. Reche C, Viana M, Rivas I, Pandolfi M, Amato F, et al. (2014) Outdoor and indoor UFP in primary schools across Barcelona. Science Total Environ 493: 943–953. doi: 10.1016/j.scitotenv.2014.06.072 PMID: 25003584
23. Reche C, Viana M, Rivas I, Bouso L, Àlvarez-Pedrerol M, et al. (2014) Child exposure to indoor and outdoor air pollutants in schools in Barcelona, Spain. Environ Int 69: 200–212. doi: 10.1016/j.envint.2014.04.009 PMID: 24875803
24. Ministry of Public Works (2012) Atlas of urban vulnerability in Spain: methodology and contents. Madrid: Ministry of Public Works.
25. de Onis M, Garza C, Onyango AW, Rolland-Cachera MF (2009) WHO development standards for infants and young children. Arch Pediatr 16: 47–53. doi: 10.1016/j.arcped.2008.10.010 PMID: 19036567
26. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry 40: 1337–1345. PMID: 11699809
27. Basner M, Babisch W, Davis A, Brink M, Clark C, et al. (2014) Auditory and non-auditory effects of noise on health. Lancet 383: 1325–1332. doi: 10.1016/S0140-6736(13)61613-X PMID: 24183105
28. Textor J, Hardt J, Knuppel S (2001) DAGitty: a graphical tool for analyzing causal diagrams. Epidemiology 22: 745.
29. Gauderman WJ, Avol E, Gilliland F, Vora H, Thomas D, et al. (2004) The effect of air pollution on lung development from 10 to 18 years of age. N Engl J Med 351: 1057–1067. PMID: 15356303
30. Calderón-Garcidueñas L, Solt AC, Henríquez-Roldán C, Torres-Jardón R, Nuse B, et al. (2008) Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response,

disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid beta-42 and alpha-synuclein in children and young adults. Toxicol Pathol 36: 289–310. doi: 10.1177/0192623307313011 PMID: 18349428

31. Vrijheid M, Martinez D, Aguilera I, Ballester F, Basterrechea M, et al. (2012) Socioeconomic status and exposure to multiple environmental pollutants during pregnancy: evidence for environmental inequity? J Epidemiol Community Health 66: 106–113. doi: 10.1136/jech.2010.117408 PMID: 20974841

32. van Kempen E, Fischer P, Janssen N, Houthuijs D, van Kamp I, et al. (2012) Neurobehavioral effects of exposure to traffic-related air pollution and transportation noise in primary schoolchildren. Environ Res 115: 18–25. doi: 10.1016/j.envres.2012.03.002 PMID: 22483436

33. Clark C, Crombie R, Head J, van Kamp I, van Kempen E, et al. (2012) Does traffic-related air pollution explain associations of aircraft and road traffic noise exposure on children health and cognition? A secondary analysis of the United Kingdom sample from the RANCH project. Am J Epidemiol 176: 327–337. doi: 10.1093/aje/kws012 PMID: 22842719

34. Mohai P, Kweon BS, Lee S, Ard K (2011) Air pollution around schools is linked to poorer student health and academic performance. Health Aff (Millwood) 30: 852–862. doi: 10.1377/hlthaff.2011.0077 PMID: 21543420

35. Guxens M, Sunyer J (2012) A review of epidemiological studies on neuropsychological effects of air pollution. Swiss Med Wkly 141: w13322. doi: 10.4414/smw.2011.13322 PMID: 22252905

36. Viana M, Rivas I, Querol X, Alastuey A, Sunyer J, et al. (2014) Indoor/outdoor relationships of quasi-ultrafine, accumulation and coarse mode particles in school environments: chemical composition and sources. Atmos Chem Phys Discuss 14: 4459–4472.

37. Block ML, Elder A, Auten RL, Bilbo SD, Chen H, et al. (2012) The outdoor air pollution and brain health workshop. Neurotoxicology 33: 972–984. doi: 10.1016/j.neuro.2012.08.014 PMID: 22981845

38. Morgan TE, Davis DA, Iwata N, Tanner JA, Snyder D, et al. (2011) Glutamatergic neurons in rodent models respond to nanoscale particulate urban air pollutants in vivo and in vitro. Environ Health Perspect 119: 1003–1009. doi: 10.1289/ehp.1002973 PMID: 21724521

39. Curtis JT, Hood AN, Chen Y, Cobb GP, Wallace DR (2010) Chronic metals ingestion by prairie voles produces sex-specific deficits in social behavior: an animal model of autism. Behav Brain Res 213: 42–49. doi: 10.1016/j.bbr.2010.04.028 PMID: 20433873

40. Biederman J, Faraone SV (2005) Attention deficit hyperactivity disorder. Lancet 366: 237–248. PMID: 16023516

41. Turkheimer E, Haley A, Waldron M, D'Onofrio B, Gottesman II (2003) Socioeconomic status modifies heritability of IQ in young children. Psychol Sci 14: 623–628. PMID: 14629696

42. Hornung RW, Lanphear BP, Dietrich KN (2009) Age of greatest susceptibility to childhood lead exposure: a new statistical approach. Environ Health Perspect 117: 1309–1312. doi: 10.1289/ehp.0800426 PMID: 19672413

Editors' Summary

Background

Human brain development is a complex and lengthy process. During pregnancy, the basic structures of the brain are formed, and the neural circuits that will eventually control movement, speech, memory, and other cognitive (thinking) functions, as well as the function of many organs, begin to be established. By the time of birth, the brain is about a quarter of its adult size, and the neural circuits that control vital bodily functions such as breathing are well developed. By contrast, the cerebral cortex—the brain region that is involved in thought and action—is poorly developed. Much of the development of the cerebral cortex happens during the first two years of life. For example, babies usually learn to crawl at about nine months. Other aspects of brain function take longer to develop. Thus, the cognitive functions that are essential for learning undergo considerable development between the ages of 6 and 10 years, and further brain changes occur during adolescence.

Why Was This Study Done?

Exposure to the air pollutants produced by the combustion of fossil fuels by vehicles during pregnancy or infancy has been associated with delays in cognitive development. Moreover, experiments in animals suggest that traffic-related air pollution is a developmental neurotoxicant—a factor that disrupts brain development. However, although many schools are located next to busy roads and although traffic-related air pollution levels peak during school hours, it is not known whether exposure of school-age children to traffic-related air pollutants impairs their cognitive development and thus their ability to learn. Here, in a prospective cohort study (the BREATHE study), the researchers assess whether exposure of children aged 7–10 years to traffic-related air pollutants in schools in Barcelona, Spain, is associated with impaired cognitive development. A prospective cohort study is an observational investigation that studies groups (cohorts) of individuals who differ with respect to a specific factor to determine how exposure to this factor affects specific outcomes.

What Did the Researchers Do and Find?

The researchers used computerized tests to measure the development of working memory (the system that holds multiple pieces of transitory information in the mind where they can be manipulated), superior working memory (working memory that involves continuous updating of the working memory buffer), and attentiveness every three months over a 12-month period in 2,715 primary school children attending 39 schools exposed to high or low levels of traffic-related air pollution and paired by socioeconomic index. That is, the researchers compared three cognitive development outcomes in the children attending each school where exposure to air pollution was high with the same outcomes in children attending a school with a similar socioeconomic index where exposure to pollution was low; school pairing was undertaken to avoid "confounding" by social class, a factor that is known to affect cognitive development. Statistical analyses of these data indicated that the increase in cognitive development over time among children attending highly polluted schools was less than that among children attending paired lowly polluted schools, even after adjusting for additional factors that affect cognitive development. Thus, for example, there was an 11.5% 12-month increase in working memory at the lowly polluted schools

but only a 7.4% 12-month increase in working memory at the highly polluted schools. Other analyses indicated that children attending schools with higher levels of traffic-related air pollutants in either the courtyard or in the classroom experienced a substantially smaller increase over the 12-month study in all three cognitive measurements than those attending schools with lower levels of pollutants.

What Do These Findings Mean?

These findings suggest that, compared with attendance at schools exposed to low levels of traffic-related air pollution, attendance at schools exposed to high levels of traffic-related air pollution is associated with a smaller increase in cognitive development over a 12-month period among 7- to 10-year-old children in Barcelona. The accuracy of these findings may be limited by residual confounding. That is, the children attending schools where traffic-related pollution is high might have shared other unknown characteristics that affected their cognitive development. Importantly, these findings do not prove that traffic-related air pollution causes impairment of cognitive development. Rather, they suggest that the developing brain may be vulnerable to traffic-related air pollution well into middle childhood, a conclusion that has implications for the design of air pollution regulations and for the location of new schools.

Additional Information

Please access these websites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.1001792.

- The US Centers for Disease Control and Prevention provides information about child development, including information about middle childhood (in English and Spanish)
- Tox Town is an interactive site that provides information about toxic chemicals and environmental health risks, including air pollution
- The US Environmental Protection Agency provides information about air pollution near roadways and health
- Wikipedia has pages on cognitive development and on air pollution (note that Wikipedia is a free online encyclopedia that anyone can edit; available in several languages)
- MedlinePlus provides links to additional resources about air pollution (in English and Spanish)
- The Centre for Research in Environmental Epidemiology website provides more information about the BREATHE study

III. Neurodevelopmental deceleration by urban fine particles from different emisison sources: a longitudinal observational study

A Section 508–conformant HTML version of this article is available at http://dx.doi.org/10.1289/EHP209.

Neurodevelopmental Deceleration by Urban Fine Particles from Different Emission Sources: A Longitudinal Observational Study

Xavier Basagaña,[1,2,3] Mikel Esnaola,[1,2,3] Ioar Rivas,[1,2,3,4] Fulvio Amato,[4] Mar Alvarez-Pedrerol,[1,2,3] Joan Forns,[1,2,3,5] Mònica López-Vicente,[1,2,3] Jesús Pujol,[6,7] Mark Nieuwenhuijsen,[1,2,3] Xavier Querol,[4] and Jordi Sunyer[1,2,3,8]

[1]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Catalonia, Spain; [2]Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain; [3]CIBER Epidemiología y Salud Pública (CIBERESP), Spain; [4]Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Catalonia, Spain; [5]Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; [6]MRI Research Unit, CRC Mar, Hospital del Mar, Barcelona, Spain; [7]Centro Investigación Biomédica en Red de Salud Mental (CIBERSAM G21), Barcelona, Spain; [8]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Catalonia, Spain

Background: A few studies have reported associations between traffic-related air pollution exposure at schools and cognitive development. The role of PM components or sources other than traffic on cognitive development has been little explored.

Objectives: We aimed to explore the role of PM sources in school air on cognitive development.

Methods: A cohort of 2,618 schoolchildren (average age, 8.5 years) belonging to 39 schools in Barcelona (Spain) was followed up for a year. Children completed computerized tests assessing working memory, superior working memory, and inattentiveness during four visits. Particulate matter ≤ 2.5 μm ($PM_{2.5}$) was measured during two 1-week campaigns in each school, both outdoors and in the classroom. Source apportionment resulted in nine sources: mineral, organic/textile/chalk, traffic, secondary sulfate and organics, secondary nitrate, road dust, metallurgy, sea spray, and heavy oil combustion. Differences in cognitive growth trajectories were assessed with mixed models with age-by-source interaction terms.

Results: An interquartile range increase in indoor traffic-related $PM_{2.5}$ was associated with reductions in cognitive growth equivalent to 22% (95% CI: 2%, 42%) of the annual change in working memory, 30% (95% CI: 6%, 54%) of the annual change in superior working memory, and 11% (95% CI: 0%, 22%) of the annual change in the inattentiveness scale. None of the other $PM_{2.5}$ sources was associated with adverse effects on cognitive development.

Conclusions: Traffic was the only source of fine particles associated with a reduction in cognitive development. Reducing air pollution from traffic at primary schools may result in beneficial effects on cognition.

Citation: Basagaña X, Esnaola M, Rivas I, Amato F, Alvarez-Pedrerol M, Forns J, López-Vicente M, Pujol J, Nieuwenhuijsen M, Querol X, Sunyer J. 2016. Neurodevelopmental deceleration by urban fine particles from different emission sources: a longitudinal observational study. Environ Health Perspect 124:1630–1636; http://dx.doi.org/10.1289/EHP209

Introduction

Particulate matter (PM) air pollution is known to produce adverse health effects with important consequences at the population level (Brunekreef and Holgate 2002; Lim et al. 2012; WHO 2013). Although the most well-established evidence for a deleterious role of PM concerns cardiovascular and respiratory diseases, emerging evidence suggests that PM exposure can also affect neurodevelopment and cognitive function (Block et al. 2012; WHO 2013). This is supported by animal studies showing neuroinflammation and neuropathological damage in the brain as well as alterations in learning and memory functions in response to air pollution exposure (Block et al. 2012).

Particulate matter is a complex mixture of different components originating from different sources. A better understanding of which components and sources of PM are responsible for the health effects is very important from the regulatory point of view. Using chemical speciation of PM measurements and source apportionment techniques, it is now possible to estimate the concentration attributable to different sources (e.g., traffic, biomass burning, industry, or natural sources) (Viana et al. 2008). Recent studies have examined the role of source-specific pollution on health outcomes, mostly cardiovascular and respiratory mortality or hospital admissions. Most of the evidence for harmful effects of air pollution refers to traffic-related air pollution, although the effects of other sources such coal combustion, shipping, road dust, or desert dust have also been documented (Cassee et al. 2013; Ostro et al. 2011; WHO 2013). The biological mechanisms leading to neurodevelopment effects may be different from those described for cardiovascular and respiratory effects, and the chemical composition of the particles, their size, or their surface area can play a relevant role. For example, suggested mechanisms include disruptions of the nasal and olfactory barrier and the blood–brain barrier allowing direct access of ultrafine particles to the brain. PM was seen in olfactory bulb neurons in children autopsies and PM-associated metals such as nickel (Ni) or vanadium (V) were detected in the brains of dogs (Calderón-Garcidueñas et al. 2003, 2008).

In a recent longitudinal study of schoolchildren, we reported that cognitive development over 1 year showed a slower increase among children attending schools with high traffic-related air pollution levels compared with children in less polluted schools (Sunyer et al. 2015). In that study, the air pollution markers used were nitrogen dioxide (NO_2), elemental carbon, and ultrafine particle number. Interestingly, fine particle (PM with aerodynamic diamater ≤ 2.5 μm; $PM_{2.5}$) mass concentrations at the studied schools were

Address correspondence to X. Basagaña, Centre for Research in Environmental Epidemiology (CREAL), c/ Doctor Aiguader 88, 08003 Barcelona, Spain. Telephone: (34) 932147306. E-mail: xbasagana@creal.cat

Supplemental Material is available online (http://dx.doi.org/10.1289/EHP209).

We acknowledge all the families participating in the study for their altruism and particularly to the schools Antoni Brusi, Baloo, Betània–Patmos, Centre d'estudis Montseny, Col.legi Shalom, Costa i Llobera, El sagrer, Els Llorers, Escola Pia de Sarrià, Escola Pia Balmes, Escola concertada Ramon Llull, Escola Lourdes, Escola Tècnica Professional del Clot, Ferran i Clua, Francesc Macià, Frederic Mistral, Infant Jesús, Joan Maragall, Jovellanos, La Llacuna del Poblenou, Lloret, Menéndez Pidal, Nuestra Señora del Rosario, Miralletes, Ramon Llull, Rius i Taulet, Pau Vila, Pere Vila, Pi d'en Xandri, Projecte, Prosperitat, Sant Ramon Nonat-Sagrat Cor, Santa Anna, Sant Gregori, Sagrat Cor Diputació, Tres Pins, Tomàs Moro, Torrent d'en Melis, and Virolai. We also acknowledge the ESCAPE (European Study of Cohorts for Air Pollution Effects) project for the design and supervision modeling of air pollution. Furthermore, X. Mayoral did the technical development of the *n*-back test; C. Persavento, J. Gonzalez, L. Bouso, and P. Figueras contributed to the field work.

This work was supported by the European Research Council (ERC; grant no. 268479: the BREATHE project). F.A. is beneficiary of the Juan de la Cierva postdoctoral grant (JCI-2012-13473) from the Spanish Ministry of Economy and Competitiveness.

The authors declare they have no actual or potential competing financial interests.

Received: 10 March 2015; Revised: 30 September 2015; Accepted: 12 April 2016; Published: 29 April 2016.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

not correlated with traffic pollution, and most of the contribution to $PM_{2.5}$ levels was due to mineral and organic sources (Amato et al. 2014). $PM_{2.5}$ levels are the universal indicator of air quality because of their overwhelming adverse association with many health indicators (WHO 2013). Here, we aim to explore the role of all the different sources of $PM_{2.5}$ in school air on cognitive development.

Methods

Design and Population

A cohort of schoolchildren in Barcelona (Spain) was followed up for a year (study period: January 2012–March 2013) as part of the BREATHE (BRain dEvelopment and Air polluTion ultrafine particles in scHool childrEn) project. Cluster sampling was performed by first selecting 40 schools and then inviting all students without special needs in grades 2 through 4 (7–10 years of age) to participate. Using a map of NO_2 levels in the city, pairs of one high-pollution and one low-pollution school matched by socioeconomic vulnerability index (census tract–level indicator based on level of education, unemployment, and occupation) and type of school (i.e., public/private) were selected. A total of 39 schools (18 pairs and 1 trio) agreed to participate and were included in the study (Sunyer et al. 2015). Participating schools were similar to the remaining schools in Barcelona in terms of the socioeconomic vulnerability index (Sunyer et al. 2015). Families of 2,897 children (59% of those eligible) agreed to participate in the study. All parents or guardians signed the informed consent form approved by the Clinical Research Ethical Committee (No. 2010/41221/I) of the IMIM-Parc de Salut MAR, Barcelona, Spain.

Outcomes: Cognitive Development

Cognitive development was assessed through long-term change in working memory and attention, because these functions grow steadily during pre-adolescence (Anderson 2002; Rueda et al. 2005). Children were evaluated every 3 months over four repeated visits using computerized tests. The computerized versions chosen [the *n*-back task on working memory (Anderson 2002) and the attentional network task (ANT) (Rueda et al. 2004)] were validated with brain imaging (Rueda et al. 2004; Thomason et al. 2009) and in the general population (Forns et al. 2014).

Briefly, in the *n*-back task, subjects are presented a sequence of stimuli in the screen (e.g., a number), one at a time, and they need to respond (i.e., hit a button) only when the current stimulus matches the one presented *n* steps before. In the present study, we analyzed only 2-back task as a measure of working memory and 3-back task as a measure of superior working memory, and used only the numbers stimuli, although other tests were also administered. These choices were based on good properties observed for these tests in the same cohort (e.g., clear age-dependent slope and little learning effect) (Sunyer et al. 2015). For each of these two tests, we measured detectability (d prime, d'), which is the normalized proportion of correctly identified targets minus the normalized proportion of false alarm hits, $d' = (z \text{ hit rate} - z \text{ false alarm rate})$. A higher d' indicates more accurate test performance. In ANT, subjects have to respond whether the central fish in a row is pointing to the left or right by pressing the corresponding button on the mouse. We used hit reaction time standard error (HRT-SE), a measure of response speed consistency, throughout the test (Sunyer et al. 2015). A high HRT-SE indicates highly variable reactions and is considered a measure of inattentiveness.

Air Pollution Exposure

Air pollution measurements were taken simultaneously for each pair of schools during two 1-week periods separated by 6 months (sampling campaign 1: January–June 2012; sampling campaign 2: September 2012–February 2013). Only a pair of schools was measured each week. High-volume samplers (MCV SA, Barcelona, Spain) for particulate matter < 2.5 μm ($PM_{2.5}$) were installed indoors in a classroom and outdoors in the playground during school hours (0900–1700 hours) from Monday through Thursday. A detailed description of the measurement campaigns and the instruments can be found elsewhere (Amato et al. 2014). Briefly, filters from samplers were divided in different pieces to determine concentrations of major and trace elements via inductively coupled plasma mass spectrometry and atomic emission spectrometry (ICP-MS and ICP-AES); concentrations of sulfate, nitrate, and chloride ions via ion chromatography (IC) and ammonium via a specific electrode; and concentrations of organic carbon (OC) and elemental carbon (EC) via a thermal-optical transmission technique (TOT).

All measurements (including indoor and outdoor measurements) were pooled to conduct the source apportionment analysis, because this was shown to provide the best results in these data (Amato et al. 2014). Source apportionment was performed using a constrained positive matrix factorization (PMF) model based on 33 chemical species. PMF is a weighted least-squares technique that allows accounting for the uncertainty associated with the analytical procedure, and was run by means of the Multilinear Engine program, which allowed the handling of *a priori* information such as the source profile of local road dust and sea spray (Amato et al. 2014). This technique returned a solution that identified nine main factors/sources responsible for the variability of $PM_{2.5}$ mass concentrations with an R^2 of 0.95. The nine sources were identified as mineral, organic/textile/chalk, traffic (that included exhaust and non-exhaust contributions), secondary sulfate and organics, secondary nitrate, road dust (resuspended street dust), metallurgy, sea spray, and heavy oil combustion (mostly from shipping in the study area and period). The elements identifying the sources are summarized in Table 1.

Outdoor and indoor long-term total and source-specific PM levels were obtained by averaging the two 1-week measures of each school. To minimize the effect of meteorology and other seasonal effects in the results, we conducted paired statistical analyses (described below) to restrict comparisons between schools that were measured simultaneously.

Contextual and Individual Covariates

Sociodemographic factors included questionnaire-based parents' responses on parental education, marital status, environmental tobacco smoke at home, and a neighborhood socioeconomic status vulnerability index (Sunyer et al. 2015) calculated both at the school and home addresses. Exposure to traffic $PM_{2.5}$ at home was estimated at the geocoded postal address using available maps based on land use regression models (Eeftens et al. 2012; Sunyer et al. 2015). Noise levels in the classroom before children arrived (as a measure of traffic-related noise) were also measured (Sunyer et al. 2015).

Statistical Analysis

Due to the multilevel nature of the data (i.e., visits within children within schools), we used linear mixed-effects models with the four repeated cognitive parameters as outcomes and random effects for child and school. Age

Table 1. Main elements identifying the estimated sources.

Source	Identifying species (tracers)
Mineral	Al, Mg, Li, Fe, Ca, Ti, Rb
Traffic	EC, Cu, Sb, Sn, Fe
Organic/textile/chalk	OC, Ca, Sr
Secondary sulfate and organics	SO_4^{2-}, NH_4^+
Secondary nitrate	NO_3^-
Road dust	Ca, Fe, Cu, Sb
Metallurgy	Zn, Pb, Cd, Mn, Cu
Sea spray	Na, Cl^-
Heavy oil combustion	V, Ni

Abbreviations: Al, aluminum; Ca, calcium; Cd, cadmium; Cl^-, chloride ion; Cu, copper; EC, elemental carbon; Fe, iron; Li, lithium; Mg, magnesium; Mn, manganese; Na, sodium; NH_4^+, ammonium cation; Ni, nickel; NO_3^-, nitrate; OC, organic carbon; Pb, lead; Rb, rubidium; Sb, antimony; Sn, tin; SO_4^{2-}, sulfate; Sr, strontium; Ti, titanium; V, vanadium; Zn, zinc.

at each visit (centered at visit 1) was included in the model to capture the growth trajectory of the cognitive test. An interaction between age and school concentrations of individual PM sources was included to capture changes in growth trajectory associated with school air pollution exposure. The latter was the effect of interest in this study. Potential confounders were identified using directed acyclic graphs (DAG) as described elsewhere (Sunyer et al. 2015), and they included sex, maternal education (primary or less/secondary/university), residential neighborhood socioeconomic status, and air pollution exposure at home. Indicators of school pair were included in the model to restrict comparisons within pairs of schools measured during the same days, thus removing potential differences in air pollution levels between schools that were attributable to meteorology or seasonality. The model equation was the following,

$$Y_{psit} = \beta_{0p} + \beta_1(\mathrm{Age}_{psit} - \mathrm{Age}_{psi1}) + \beta_2(\mathrm{PM\_source})_{ps} + \beta_3(\mathrm{Age}_{psit} - \mathrm{Age}_{psi1}) \times (\mathrm{PM\_source})_{ps} + \mathbf{Z\eta} + u_{ps} + v_{psi} + \varepsilon_{psit},$$

where Y_{psit} is the cognitive test result for subject i in school s (belonging to pair p) at visit t, $t = (1,2,3,4)$, β_{0p} are pair-specific intercepts, $\mathbf{Z}$ is a matrix including all confounders, $\boldsymbol{\eta}$ is a vector of parameters associated to confounders, u_{ps} are random effects at school level, assumed normally distributed with mean 0 and variance σ_u^2, v_{psi} are random effects associated with subject i in school s, assumed normally distributed with mean 0 and variance σ_v^2, and ε_{psit} are the model residuals assumed normally distributed with mean 0 and variance σ_e^2. Deviations from linearity were assessed with generalized additive mixed models. Analyses were repeated without the pair indicator and also with further adjustment for total $PM_{2.5}$ levels and the interaction between age and total $PM_{2.5}$ (Mostofsky et al. 2012). The interactions between age and maternal education ($p > 0.15$ for all outcomes) and age and socioeconomic status ($p > 0.5$ for all outcomes) were unrelated to cognitive development and were not included in the models.

Models included only $PM_{2.5}$ concentrations from a single source at a time, and separate models were fitted for each source. Likewise, separate models were fitted for indoor and outdoor concentrations. Regression coefficients were rescaled to represent the change in the outcome associated with an interquartile range change in source-specific $PM_{2.5}$ levels.

We also provided the results using tracers (chemical elements identifying the source) instead of sources for those $PM_{2.5}$ sources showing significant or suggestive adverse effects on cognition. Statistical significance was set at $p < 0.05$.

Results

Table 2 summarizes the characteristics of the selected schools with respect to the high versus low air pollution indicator used at the design stage. Schools with high air pollution showed lower area-level deprivation and less greenness, had a lower percentage of public schools and higher indoor noise levels, were closer to busy roads, and their students tended to live closer to the school. Education quality was equivalent in the two groups.

The average age of participants at baseline was 8.5 years. A total of 2,618 (90.3%) children had data on the three outcomes in at least one visit. Children without data on cognitive outcomes more often attended public schools (54% vs. 33%), but there were no differences in terms of school vulnerability index (0.45 vs. 0.42). Around half of the children were girls and they attended 2nd, 3rd, and 4th grade (37%, 36%, and 27%, respectively) in the first visit. Thirty-four percent of them attended a public school, and the rest attended a private school (Table 3). More than half of the mothers (58.9%) had a university education, whereas 12.5% had at most achieved primary education. Thirty-one percent of them lived in areas of high deprivation according to the socioeconomic status vulnerability index. More details of the study population can be found elsewhere (Sunyer et al. 2015).

During the 1-year follow-up, working memory increased on average by 13.0%, superior working memory by 16.5%, and inattentiveness decreased by 14% (Table 3). At baseline, lower scores were observed for girls, children attending public schools, children from mothers with low education, and children living in more deprived areas. Children from public schools showed a greater change in superior working memory over follow-up than those from private schools. Change over follow-up was not significantly associated with other characteristics in crude analyses (Table 3).

Table 2. Characteristics of selected schools according to the air pollution indicator used at the design stage (city map of NO_2 levels).

Characteristic	Low air pollution	High air pollution
Number	20	19
Socioeconomic vulnerability index	0.52 ± 0.24	0.41 ± 0.16
School greenness (NDVI)	0.31 ± 0.10	0.15 ± 0.03
Public school (%)	55	42
Education quality (PISA 2012)	3.9 ± 1.3	3.9 ± 1.8
Noise level in classroom (dB)	37.2 ± 4.9	40.1 ± 5.0
Distance to busy roads (m)	369 ± 357	118 ± 178
Average distance to children home (m)	2,432 ± 2,338	1,048 ± 1,613

Abbreviations: NDVI, Normalized Difference Vegetation Index; PISA, Programme for International Student Assessment. Data are number, percent, or mean ± SD. This table is a partial reproduction of published work (Table 5 in Sunyer et al. 2015).

Table 3. Mean (± SD) of cognitive outcomes by characteristics of participants.

		Working memory (WM) (2-back numbers, $d' \times 100$)		Superior WM (3-back numbers, $d' \times 100$)		Inattentiveness (HRT-SE, ms)	
Characteristics	n (%)[a]	Baseline	Change	Baseline	Change	Baseline	Change
All	2,618 (100)	224 ± 126	30 ± 156	118 ± 100	20 ± 130	272 ± 90	−38 ± 89
Sex							
Male	1,316 (50.3)	229 ± 129	25 ± 155	123 ± 103	16 ± 129	261 ± 89*	−35 ± 89
Female	1,302 (49.7)	220 ± 122	35 ± 157	113 ± 96	23 ± 132	284 ± 89	−41 ± 88
Type of school							
Public	860 (32.8)	215 ± 129*	36 ± 149	111 ± 102*	32 ± 126*	274 ± 91	−39 ± 89
Private	1,758 (67.2)	229 ± 124	27 ± 159	121 ± 99	14 ± 132	271 ± 89	−38 ± 89
Maternal education							
Primary or less	337 (12.9)	188 ± 134*	25 ± 162	82 ± 91*	25 ± 130	308 ± 88*	−40 ± 92
Secondary	743 (28.4)	213 ± 124	32 ± 160	118 ± 102	14 ± 132	274 ± 90	−31 ± 93
University	1,538 (58.7)	237 ± 122	30 ± 153	126 ± 100	21 ± 130	264 ± 88	−41 ± 85
SES vulnerability at home							
Less deprived	980 (37.4)	233 ± 124*	29 ± 151	127 ± 102*	18 ± 131	265 ± 89*	−38 ± 87
Middle deprived	807 (30.8)	227 ± 125	27 ± 159	116 ± 98	27 ± 131	274 ± 92	−37 ± 91
High deprived	831 (31.7)	212 ± 128	34 ± 159	110 ± 100	15 ± 130	281 ± 88	−39 ± 89
School Pair							
Low polluted	1,328	226 ± 125	36 ± 154	120 ± 100	23 ± 131	272 ± 89	−42 ± 86
High polluted	1,290	222 ± 126	24 ± 158	116 ± 100	16 ± 130	273 ± 90	−34 ± 91
Residential $PM_{2.5}$ from traffic							
1st quartile	635	224 ± 128	32 ± 157	121 ± 101	17 ± 125	275 ± 94	−38 ± 83
2nd quartile	662	229 ± 124	22 ± 159	117 ± 103	21 ± 139	272 ± 90	−38 ± 86
3rd quartile	659	224 ± 126	37 ± 150	123 ± 98	15 ± 128	271 ± 89	−41 ± 93
4th quartile	662	220 ± 125	29 ± 157	111 ± 98	25 ± 129	271 ± 86	−37 ± 92

[a]Number of participants with data at baseline. *$p < 0.05$ when testing equality between groups.

The median of school-averaged $PM_{2.5}$ mass concentrations was 28 μg/m^3 outdoors and 36 μg/m^3 indoors. Mineral (27%) was the source contributing the highest concentration to outdoor $PM_{2.5}$ levels, followed by traffic (17%), organic/textile/chalk (16%), sulfate (14%), nitrate (13%), and smaller contributions of road dust (4%), metallurgy (4%), sea salt (3%), and heavy oil combustion (2%) (Figure 1). Indoor concentrations were in general smaller and followed the same ordering than outdoor sources, except for organic/textile/chalk, with a strong indoor origin and representing the highest contribution to indoor $PM_{2.5}$ (45%). Table 4 describes the variation of concentrations by source across schools. Mineral exhibited the largest variation, with schools below the 25th percentile having mineral concentrations that represented at most 5.5% of the total levels, whereas in schools above the 75th percentile mineral contributed > 34% of the total $PM_{2.5}$ levels. The indoor organic/textile/chalk source also showed large variations between schools. The interquartile range for the traffic source was higher for indoor than for outdoor levels, probably reflecting the effect of class orientation on infiltration (Amato et al. 2014). More details on source apportionment results can be found elsewhere (Amato et al. 2014).

Correlations between indoor and outdoor levels of the same sources were generally greater than 0.7 (see Table S1). Exceptions to this pattern were organic/textile/chalk, with indoor–outdoor correlation close to zero; road dust, with a correlation of 0.14; and to a lesser extent mineral, with a correlation of 0.64. The highest correlations between school levels of outdoor sources were in the 0.5–0.6 range, including the correlations of mineral with organic/textile/chalk, sea salt, and road dust, and the pairs metallurgy–secondary nitrate, heavy oil combustion–secondary sulfate, and road dust–organic/textile/chalk. With regard to indoor sources, the pairs of sources mentioned above for outdoor levels showed

Table 4. Description of source contributions to $PM_{2.5}$ in terms of mass and as a percentage of total $PM_{2.5}$ mass.

	Concentration (μg/m^3)				Percent			
	Percentile				Percentile			
Source	25th	50th	75th	IQR	25th	50th	75th	IQR
Outdoor $PM_{2.5}$	22.6	28.1	35.8	13.2	—	—	—	—
Mineral	1.2	2.6	12.7	11.5	5.5	11.3	33.7	28.2
Traffic	4.1	5.2	6.8	2.7	12.9	20.5	26.1	13.3
Organic/textile/chalk	2.3	4.8	7.1	4.7	10.8	14.7	20.2	9.4
Secondary sulfate and organics	2.6	4.5	5.7	3.1	10.3	13.7	24.7	14.4
Secondary nitrate	1.9	3.2	5.1	3.2	7.1	11.3	15.6	8.5
Road dust	0.6	1.1	1.8	1.2	2.3	4.3	5.5	3.2
Metallurgy	0.9	1.2	1.5	0.6	3.1	3.9	5.4	2.3
Sea spray	0.5	0.7	1.1	0.6	1.8	2.3	3.5	1.7
Heavy oil combustion	0.5	0.6	0.8	0.4	1.6	2.2	2.8	1.1
Indoor $PM_{2.5}$	29.2	35.6	41.5	12.3	—	—	—	—
Mineral	2.0	3.9	7.2	5.2	6.5	11.6	20.6	14.1
Traffic	3.0	4.4	6.8	3.8	9.7	12.7	20.0	10.3
Organic/textile/chalk	12.3	15.3	20.1	7.8	37.2	44.8	48.9	11.6
Secondary sulfate and organics	2.3	3.6	5.4	3.1	7.2	10.7	15.0	7.8
Secondary nitrate	0.9	1.1	1.9	1.0	2.6	3.8	5.3	2.8
Road dust	0.5	1.3	2.1	1.6	1.7	3.3	5.2	3.5
Metallurgy	0.7	0.9	1.2	0.5	2.0	2.8	3.3	1.3
Sea spray	0.6	0.7	1.0	0.5	1.7	2.0	2.7	1.1
Heavy oil combustion	0.4	0.6	0.7	0.3	1.1	1.4	2.0	0.9

IQR, interquartile range.

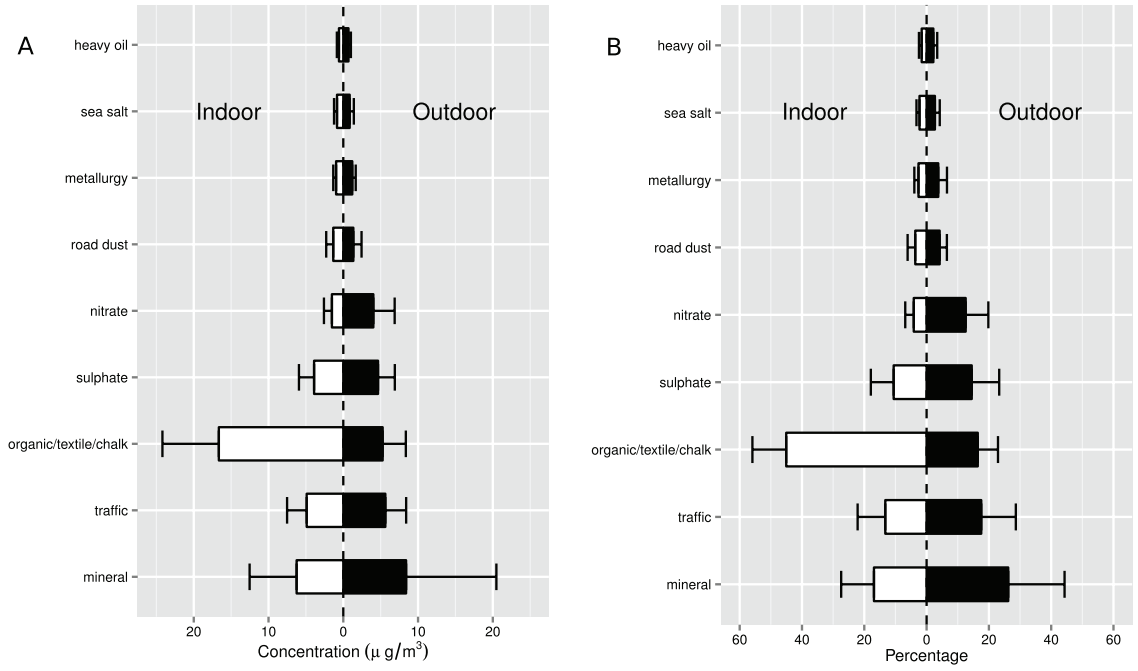

Figure 1. Average source concentrations (*A*) and percent of $PM_{2.5}$ concentrations (*B*) inside (indoor) and outside (outdoor) of schools. Error bars indicate mean ± SD.

similarly high correlations, whereas secondary nitrate also showed high correlations with secondary sulfate and heavy oil combustion, and traffic showed a negative correlation of –0.52 with road dust (see Table S2).

Figure 2 (see also Table S3) displays the change in cognitive outcomes over the follow-up period for an interquartile range increase in source-specific $PM_{2.5}$ concentrations. Results in unadjusted analyses were fairly similar to adjusted ones (see Table S3). Changes from the first to the third quartile in the indoor traffic source were associated with a significant reduction in working memory of –5.6 [95% confidence interval (CI): –10.7, –0.5], equivalent to 22% of the annual change experienced by the participants (see Table S3); a reduction of superior working memory of –5.1 (95% CI: –9.2, –1.1), equivalent to 30% of the annual change; and an increase of 3.6 (95% CI: 0.0, 7.1) in inattentiveness scale, equivalent to 11% of the annual change. Associations were smaller for outdoor concentrations of traffic $PM_{2.5}$, although results were still significant for superior working memory and inattentiveness. No significant associations were found for $PM_{2.5}$ mass concentrations from other sources, except for a positive association for outdoor concentrations of $PM_{2.5}$ from mineral origin and superior working memory. Outdoor levels of $PM_{2.5}$ from heavy oil combustion showed deleterious effects on inattentiveness and working memory (p = 0.05 and 0.09, respectively). No important deviations from linearity were detected (data not shown). When analyses were repeated excluding the school pair indicator or adjusting by environmental tobacco smoke exposure at home or traffic-related noise at school results were almost the same (data not shown). Further adjustment for total $PM_{2.5}$ levels produced only minimal changes in the results (see Table S4).

Figure S1 provides the results when using the tracers for traffic, organic/textile/chalk, secondary nitrate, and heavy oil combustion. EC was significantly associated with all outcomes, while copper (Cu) and antimony (Sb) showed significant associations only for inattentiveness. No significant associations were found for tin (Sn), iron (Fe), nitrate (NO_3^-), V, Ni, OC, calcium (Ca) or strontium (Sr).

Discussion

In a longitudinal study assessing cognitive development of schoolchildren during 1 year, we found that children attending schools with high levels of traffic-related $PM_{2.5}$ showed a slower cognitive development. None of the other $PM_{2.5}$ sources (mineral, organic/textile/chalk, sulfate, nitrate, road dust, metallurgy, and sea spray) showed a deleterious association with cognitive development, although associations for heavy oil combustion were also suggested. Associations with traffic pollution were stronger when considering indoor levels and these associations were detected for working memory, superior working memory, and inattentiveness. These results suggest that fine particles from traffic may produce neurotoxic effects, and that exposure to such particles at primary schools can result in a deceleration of cognitive development.

Our previous study was the first to relate primary school levels of air pollution to cognitive development (growth) in a longitudinal setting (Sunyer et al. 2015). A few other studies have related air pollution exposure at schools to neurobehavioral function at a single point in time (van Kempen et al. 2012; Wang et al. 2009), although others found no association (Clark et al. 2012). Other cross-sectional studies related personal or residential air pollution exposure with cognitive outcomes, and most of them reported positive associations (Chiu et al. 2013; Franco Suglia et al. 2008; Grahame et al. 2014; Guxens et al. 2014; Guxens and Sunyer 2012; Perera et al. 2012). Our study is the first to perform source apportionment of PM and examine the relationship of each individual source with cognitive development in children. In our study, PM from traffic was the only source associated with a slower cognitive development, which agrees with our previously published result

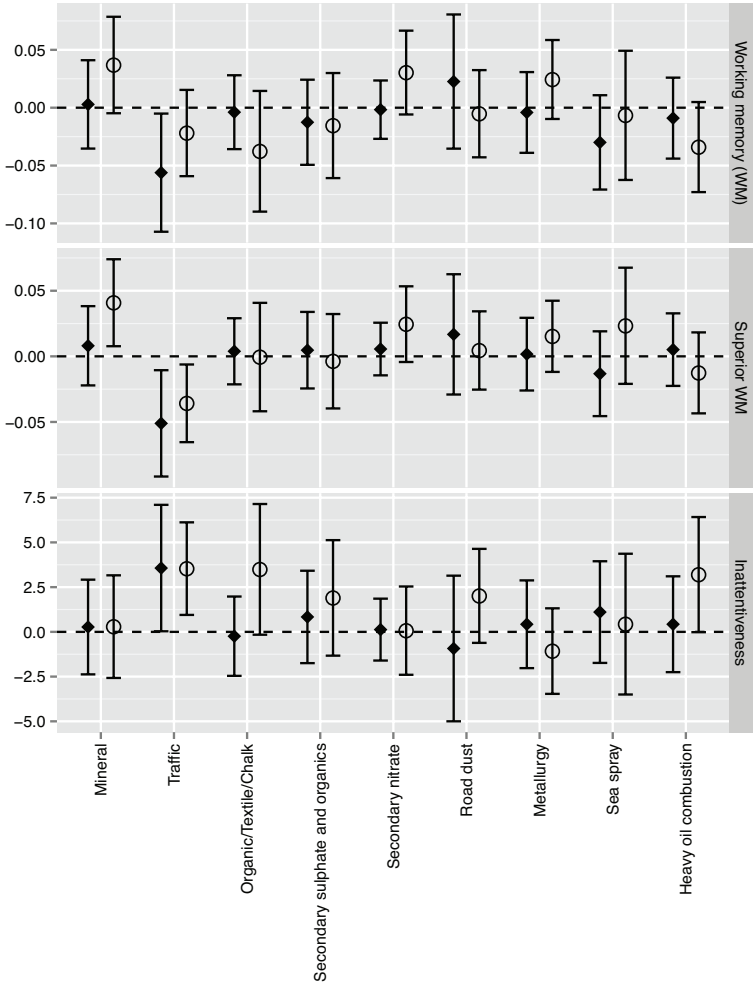

Figure 2. Change (95% CI) in cognitive growth per interquartile range increase in school source-specific $PM_{2.5}$ mass concentrations. Models were adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, residential $PM_{2.5}$ levels from traffic and school pair; school and subject included as nested random effects. Working memory measured with 2-back Numbers, $d' \times 100$. Superior working memory measured with 3-back numbers, $d' \times 100$. Inattentiveness measures with HRT-SE, ms. Black diamonds (♦): indoor concentrations; open circles (o): outdoor concentrations.

on the effects of EC in this same cohort (Sunyer et al. 2015). Although the traffic source includes also non-exhaust particles, the correlation of the source with EC was 0.89. Most of the previous studies used markers of traffic air pollution such as EC or black carbon (BC), NO_2, $PM_{2.5}$ absorbance, or polycyclic aromatic hydrocarbons (PAHs) (Chiu et al. 2013; Franco Suglia et al. 2008; Grahame et al. 2014; Guxens et al. 2014; Guxens and Sunyer 2012; Perera et al. 2012).

The role of PM sources other than traffic on cognitive development has been little explored, although some studies exist on industrial pollution. An ecological study in Michigan (USA) found an increased percentage of school failure in schools with higher levels of industrial pollution (Mohai et al. 2011). Other studies in children have found that manganese (Mn) concentrations from mining and industry were associated with impaired verbal intellectual function and motor skills (Lucchini et al. 2012). In our study area, air pollution from industry sources was low, which may be the reason why we did not find associations with this source. Besides, industry emissions depend strongly on industry type, so associations are expected to vary by study setting. Our results in relation to heavy oil combustion were inconclusive. We found an association of outdoor levels of heavy oil combustion with two of the outcomes, but chance could not be excluded due to the *p*-values at the limit of significance and the lack of association with indoor levels or with Ni and V, the main elements defining this source. Unexpectedly, we found that exposure to mineral particles was beneficial for superior working memory. This finding was not observed for indoor levels and we do not know of other studies that investigated the link between mineral particles and cognitive development. This result could be a chance finding. Schools with higher mineral concentrations had sandy playgrounds, so an alternative explanation is that they also have more greenness, which can have beneficial effects on cognitive development (Dadvand et al. 2015). However, further adjustment for greenness did not change the mineral results (data not shown).

When examining the effects of chemical elements (tracers of sources) on cognitive development, the most consistent results were found for EC. Cu and Sb were significantly associated only with inattentiveness. There is still debate on which specific components linked to traffic produce health effects, but there seems to be some consensus in that the health effects are not produced by EC alone, but by other co-emissions such as semi-volatile organic compounds (SVOCs) and PAHs that are adsorbed onto the EC core (Grahame et al. 2014). In our study, we also found some suggestions for deleterious effects of particles from heavy oil combustion, which goes along with the hypothesis of particles from combustion being harmful for the brain. Thus, our findings of slower cognitive development associated with exposure to EC may have implications beyond the effects of traffic emissions. For example, biomass burning can also be an important source of EC/BC and PAHs (Grahame et al. 2014), and the high concentrations of indoor pollution from biomass burning in developing countries could have important effects on the cognitive development of exposed children. In our study area, PM mass concentrations from biomass burning were negligible and this question could not be investigated (Reche et al. 2012).

Toxicological studies support the neurotoxic effects of motor exhaust particles (Grahame et al. 2014). The main biological mechanisms involve proinflammatory and inflammatory effects in the brain following brain deposition of particles or as a result of systemic inflammation produced by deposition of particles in the respiratory tract and alteration of blood–brain barrier function (Block et al. 2012; Calderón-Garcidueñas et al. 2008; Grahame et al. 2014). The brain may be especially vulnerable to oxidative stress, and diesel particles (highly enriched in EC) have been shown to activate microglia, which can produce neurotoxicity via oxidative stress (Block et al. 2012; Grahame et al. 2014). Fine particles from other vehicle sources such as brakes could also contribute to the effects beyond motor exhaust, given the association observed for elements generated by brakes abrasion with inattentiveness in our study and their established potential neurotoxicity (Bandmann et al. 2015). Ni and V, which could also lead to oxidative stress, were not associated with cognition in our study.

Our study had several strengths, including its longitudinal design with repeated outcome measurements and the direct measurements of air pollution both indoors and outdoors at schools. The study also had some limitations, such as a relatively small number of schools and the possibility of residual confounding by socioeconomic characteristics. The latter was extensively explored in our previous study, and all analyses suggested the observed effects were not attributable to residual confounding (Sunyer et al. 2015). In our analyses, schools were matched by socioeconomic characteristics and type of school, thus reducing potential differences, and although children from more educated families attended schools with lower air pollution levels, differences were small. Because of the observational nature of the study, it cannot be ruled out that children attending schools with high levels of pollution shared other unmeasured characteristics (e.g., not captured socioeconomic dimensions, different level of social interaction) that affected their cognitive development. Differences in cognition were already present at the beginning of the study. This would still be consistent with our hypotheses, because the cognitive functions studied were already developing in the previous years and children were already exposed to school air pollution. Importantly, we observed that these differences widened during the study period, but we could not ignore that children of more-polluted schools were already in a slower cognitive trajectory because of early-life exposures or socioeconomic factors.

Air pollution levels were based on direct measurements at schools on two different seasons. Although this may represent an improvement over previous papers, which used models to estimate air pollution concentrations at schools, our estimations were still imperfect estimates of annual concentrations. This measurement error is unlikely to be related to school characteristics, in which case it would bias the results toward the null. Another limitation was that, in order to increase the statistical power of source apportionment, this was applied to the joint set of all indoor and outdoor measurements, which may generate some artifacts in source identification. Our data could also be affected by other issues in source estimation, such as imperfect separation of road dust and the mineral sources. Uncertainty in source estimation, if properly accounted for, would widen our confidence intervals (Kioumourtzoglou et al. 2014). We did not have data on other pollutants such as gases or volatile compounds that may be related to cognitive effects, but these are expected to be correlated with the estimated sources. Further studies in different settings are needed to assess the generalizability of these results. Finally, it is worth mentioning that we examined working memory and inattentiveness, but not other domains of cognition such as visuospatial ability or language.

Conclusions

This study aimed to investigate whether levels of PM at schools were associated with cognitive development separately for each PM source. We found that levels of PM from traffic were associated with important reductions in cognitive growth over a 1-year period in primary school children. Future studies should examine whether the effects observed at primary school age are long-lasting and have consequences over the life course.

References

Amato F, Rivas I, Viana M, Moreno T, Bouso L, Reche C, et al. 2014. Sources of indoor and outdoor $PM_{2.5}$ concentrations in primary schools. Sci Total Environ 490:757–765.

Anderson P. 2002. Assessment and development of executive function (EF) during childhood. Child Neuropsychol 8(2):71–82.

Bandmann O, Weiss KH, Kaler SG. 2015. Wilson's disease and other neurological copper disorders. Lancet Neurol 14(1):103–113.

Block ML, Elder A, Auten RL, Bilbo SD, Chen H, Chen JC, et al. 2012. The outdoor air pollution and brain health workshop. Neurotoxicology 33(5):972–984.

Brunekreef B, Holgate ST. 2002. Air pollution and health. Lancet 360(9341):1233–1242.

Calderón-Garcidueñas L, Maronpot RR, Torres-Jardon R, Henríquez-Roldán C, Schoonhoven R, Acuña-Ayala H, et al. 2003. DNA damage in nasal and brain tissues of canines exposed to air pollutants is associated with evidence of chronic brain inflammation and neurodegeneration. Toxicol Pathol 31(5):524–538.

Calderón-Garcidueñas L, Solt AC, Henríquez-Roldán C, Torres-Jardón R, Nuse B, Herritt L, et al. 2008. Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid β-42 and α-synuclein in children and young adults. Toxicol Pathol 36(2):289–310.

Cassee FR, Héroux ME, Gerlofs-Nijland ME, Kelly FJ. 2013. Particulate matter beyond mass: recent health evidence on the role of fractions, chemical constituents and sources of emission. Inhal Toxicol 25(14):802–812.

Chiu YH, Bellinger DC, Coull BA, Anderson S, Barber R, Wright RO, et al. 2013. Associations between traffic-related black carbon exposure and attention in a prospective birth cohort of urban children. Environ Health Perspect 121:859–864, doi: 10.1289/ehp.1205940.

Clark C, Crombie R, Head J, van Kamp I, van Kempen E, Stansfeld SA. 2012. Does traffic-related air pollution explain associations of aircraft and road traffic noise exposure on children's health and cognition? A secondary analysis of the United Kingdom sample from the RANCH project. Am J Epidemiol 176(4):327–337.

Dadvand P, Nieuwenhuijsen MJ, Esnaola M, Forns J, Basagaña X, Alvarez-Pedrerol M, et al. 2015. Green spaces and cognitive development in primary schoolchildren. Proc Natl Acad Sci USA 112(26):7937–7942, doi: 10.1073/pnas.1503402112.

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. 2012. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. Environ Sci Technol 46(20):11195–11205.

Forns J, Esnaola M, López-Vicente M, Suades-González E, Alvarez-Pedrerol M, Julvez J, et al. 2014. The n-back test and the attentional network task as measures of child neuropsychological development in epidemiological studies. Neuropsychology 28(4):519–529.

Franco Suglia S, Gryparis A, Schwartz J, Wright RJ. 2008. Association between traffic-related black carbon exposure and lung function among urban women. Environ Health Perspect 116:1333–1337, doi: 10.1289/ehp.11223.

Grahame TJ, Klemm R, Schlesinger RB. 2014. Public health and components of particulate matter: the changing assessment of black carbon. J Air Waste Manag Assoc 64(6):620–660.

Guxens M, Garcia-Esteban R, Giorgis-Allemand L, Forns J, Badaloni C, Ballester F, et al. 2014. Air pollution during pregnancy and childhood cognitive and psychomotor development: six European birth cohorts. Epidemiology 25(5):636–647.

Guxens M, Sunyer J. 2012. A review of epidemiological studies on neuropsychological effects of air pollution. Swiss Med Wkly 141:w13322, doi: 10.4414/smw.2011.13322.

Kioumourtzoglou MA, Coull BA, Dominici F, Koutrakis P, Schwartz J, Suh H. 2014. The impact of source contribution uncertainty on the effects of source-specific $PM_{2.5}$ on hospital admissions: a case study in Boston, MA. J Expo Sci Environ Epidemiol 24(4):365–371.

Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380(9859):2224–2260.

Lucchini RG, Guazzetti S, Zoni S, Donna F, Peter S, Zacco A, et al. 2012. Tremor, olfactory and motor changes in Italian adolescents exposed to historical ferro-manganese emission. Neurotoxicology 33(4):687–696.

Mohai P, Kweon BS, Lee S, Ard K. 2011. Air pollution around schools is linked to poorer student health and academic performance. Health Aff (Millwood) 30(5):852–862.

Mostofsky E, Schwartz J, Coull BA, Koutrakis P, Wellenius GA, Suh HH, et al. 2012. Modeling the association between particle constituents of air pollution and health outcomes. Am J Epidemiol 176(4):317–326.

Ostro B, Tobias A, Querol X, Alastuey A, Amato F, Pey J, et al. 2011. The effects of particulate matter sources on daily mortality: a case-crossover study of Barcelona, Spain. Environ Health Perspect 119:1781–1787, doi: 10.1289/ehp.1103618.

Perera FP, Tang D, Wang S, Vishnevetsky J, Zhang B, Diaz D, et al. 2012. Prenatal polycyclic aromatic hydrocarbon (PAH) exposure and child behavior at age 6–7 years. Environ Health Perspect 120:921–926, doi: 10.1289/ehp.1104315.

Reche C, Viana M, Amato F, Alastuey A, Moreno T, Hillamo R, et al. 2012. Biomass burning contributions to urban aerosols in a coastal Mediterranean city. Sci Total Environ 427–428:175–190.

Rueda MR, Fan J, McCandliss BD, Halparin JD, Gruber DB, Lercari LP, et al. 2004. Development of attentional networks in childhood. Neuropsychologia 42(8):1029–1040.

Rueda MR, Rothbart MK, McCandliss BD, Saccomanno L, Posner MI. 2005. Training, maturation, and genetic influences on the development of executive attention. Proc Natl Acad Sci USA 102(41):14931–14936.

Sunyer J, Esnaola M, Alvarez-Pedrerol M, Forns J, Rivas I, López-Vicente M, et al. 2015. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study. PLoS Med 12(3):e1001792, doi: 10.1371/journal.pmed.1001792.

Thomason ME, Race E, Burrows B, Whitfield-Gabrieli S, Glover GH, Gabrieli JD. 2009. Development of spatial and verbal working memory capacity in the human brain. J Cogn Neurosci 21(2):316–332.

van Kempen E, Fischer P, Janssen N, Houthuijs D, van Kamp I, Stansfeld S, et al. 2012. Neurobehavioral effects of exposure to traffic-related air pollution and transportation noise in primary schoolchildren. Environ Res 115:18–25.

Viana M, Kuhlbusch TAJ, Querol X, Alastuey A, Harrison RM, Hopke PK, et al. 2008. Source apportionment of particulate matter in Europe: a review of methods and results. J Aerosol Sci 39(10):827–849.

Wang S, Zhang J, Zeng X, Zeng Y, Wang S, Chen S. 2009. Association of traffic-related air pollution with children's neurobehavioral functions in Quanzhou, China. Environ Health Perspect 117:1612–1618, doi: 10.1289/ehp.0800023.

WHO (World Health Organization). 2013. Review of Evidence on Health Aspects of Air Pollution–REVIHAAP Project: Final Technical Report. Copenhagen, Denmark:WHO Regional Office for Europe. Available: http://www.euro.who.int/__data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf?ua=1 [accessed 16 August 2016].

IV. A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated RNA-seq experiments

Esnaola *et al. BMC Bioinformatics* 2013, **14**:254
http://www.biomedcentral.com/1471-2105/14/254

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated RNA-seq experiments

Mikel Esnaola[1,6], Pedro Puig[2], David Gonzalez[3], Robert Castelo[4,5*] and Juan R Gonzalez[1,2,5,6*]

Abstract

Background: High-throughput RNA sequencing (RNA-seq) offers unprecedented power to capture the real dynamics of gene expression. Experimental designs with extensive biological replication present a unique opportunity to exploit this feature and distinguish expression profiles with higher resolution. RNA-seq data analysis methods so far have been mostly applied to data sets with few replicates and their default settings try to provide the best performance under this constraint. These methods are based on two well-known count data distributions: the Poisson and the negative binomial. The way to properly calibrate them with large RNA-seq data sets is not trivial for the non-expert bioinformatics user.

Results: Here we show that expression profiles produced by extensively-replicated RNA-seq experiments lead to a rich diversity of count data distributions beyond the Poisson and the negative binomial, such as Poisson-Inverse Gaussian or Pólya-Aeppli, which can be captured by a more general family of count data distributions called the Poisson-Tweedie. The flexibility of the Poisson-Tweedie family enables a direct fitting of emerging features of large expression profiles, such as heavy-tails or zero-inflation, without the need to alter a single configuration parameter. We provide a software package for R called `tweeDEseq` implementing a new test for differential expression based on the Poisson-Tweedie family. Using simulations on synthetic and real RNA-seq data we show that `tweeDEseq` yields *P*-values that are equally or more accurate than competing methods under different configuration parameters. By surveying the tiny fraction of sex-specific gene expression changes in human lymphoblastoid cell lines, we also show that `tweeDEseq` accurately detects differentially expressed genes in a real large RNA-seq data set with improved performance and reproducibility over the previously compared methodologies. Finally, we compared the results with those obtained from microarrays in order to check for reproducibility.

Conclusions: RNA-seq data with many replicates leads to a handful of count data distributions which can be accurately estimated with the statistical model illustrated in this paper. This method provides a better fit to the underlying biological variability; this may be critical when comparing groups of RNA-seq samples with markedly different count data distributions. The `tweeDEseq` package forms part of the Bioconductor project and it is available for download at http://www.bioconductor.org.

*Correspondence: robert.castelo@upf.edu; jrgonzalez@creal.cat
[4] Department of Experimental and Health Sciences, Research Program on Biomedical Informatics (GRIB), Universitat Pompeu Fabra, Barcelona, Spain
[1] Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
Full list of author information is available at the end of the article

Background

High-throughput gene expression profiling across samples constitutes one of the primary tools for characterizing phenotypes at molecular level. One of the main advantages of the rapidly evolving massive scale cDNA sequencing assay for this purpose (RNA-seq [1]), over the hybridization-based microarray technology, is a much larger dynamic range of detection. However, the extent to which this feature is fully exploited depends entirely on the way the resulting data is analyzed when addressing a particular biological question. For instance, in the identification of genes that significantly change their expression levels between groups of samples, also known as differential expression (DE).

For DE analysis, after some pre-processing steps that include the alignment of the sequenced reads to a reference genome and their summarization into features of interest (e.g., genes), raw RNA-seq data is transformed into an initial table of counts. This table should then be normalized [2-4] in order to adjust for both technical variability and the expression properties of the samples, such that the estimated normalization factors and offsets applied to the RNA-seq count data describe as accurately as possible the relative number of copies of each feature throughout every sample. As opposed to the continuous nature of log-scale fluorescence units in microarray data, RNA-seq expression levels are defined by discrete count data, and therefore, require specific DE analysis techniques.

Detection of DE genes using RNA-seq data was firstly based on using models assuming a Poisson distribution [5] with one single parameter, the mean, which simultaneously determines the variance of the distribution. Given that the observed variation in read counts is much larger than the mean (overdispersion), researchers have proposed the use of negative binomial (NB) distributions [6-8] which are defined by two parameters: the mean and the dispersion. However, the larger power of RNA-seq to capture biological variability can potentially introduce into count data not only overdispersion, but also oddities such as zero-inflation (i.e., in lowly expressed genes, the proportion of zero counts may be greater than expected under an NB distribution) and heavy tail behavior (i.e., a large dynamic range within the same expression profile), specially when many biological replicates are available. Under these circumstances even a two-parameter NB distribution may not provide an adequate fit to the data (see Figure 1). In turn, this may lead to incorrect statistical inferences resulting in lists of DE genes with a potentially increased number of false positive calls and poor reproducibility. To overcome this problem, methods based on the NB distribution [6-11] use sophisticated moderation techniques that borrow information across genes and exploit the mean-variance relationship in count data to improve the estimation of the NB dispersion parameter. This requires, however, that the parameter configuration is calibrated for the most appropriate moderation regime which may depend on features such as sample size, the magnitude of the fold-change, the variability of expression levels, the fraction of genes undergoing differential expression and the overall expression level.

In this paper we propose to approach this problem by using other count data distributions that fit expression profiles better than the NB without the need to alter configuration parameters. The rest of the paper is organized as follows. Using a large RNA-seq data set of HapMap lymphoblastoid cell lines (LCLs) derived from $n = 69$ unrelated Nigerian (YRI) individuals [12], we start by assessing the goodness of fit of extensively replicated expression profiles to the NB distribution, showing a lack of fit for an important fraction of genes. We illustrate how a more flexible family of count-data probability distributions, called the Poisson-Tweedie, provides a better fit to these expression profiles. We provide data supporting the hypothesis that the lack of fit to NB distributions may be related to the dynamics of gene expression unveiled by RNA-seq technology. We then introduce a new test for differential expression analysis in RNA-seq data based on the Poisson-Tweedie family of distributions. We demonstrate with simulations on synthetic and real RNA-seq data how a single run of our approach provides *P*-values that are equally or more accurate than NB-based competing methods calibrated with a variety of configuration parameters. Finally, by surveying the tiny fraction of sex-specific gene expression changes in LCL samples, we approach the problem of assessing accuracy in DE analysis with real RNA-seq data and show that, in the context of extensively replicated RNA-seq experiments, `tweeDEseq` yields better performance than competing NB-based methods without the need to make an informed decision on the configuration of parameters. This improvement is reported in terms of precision and recall of DE genes and reproducibility of the significance levels with respect to matching microarray experiments.

Results and discussion

The results we provide in this paper are based on data from a previously published large RNA-seq experiment [12] and on our own simulated count data. We downloaded and pre-processed the HapMap LCL raw RNA-seq data, consisting of $n = 69$ samples from unrelated YRI individuals, with our own pipeline (see Methods). The resulting table of counts consists of 38,415 genes by 69 samples. We filtered out genes with very low expression levels and used different normalization methods [2,4] (see Methods) to ensure that the results described below do not depend on this fundamental step. In fact, we have observed that normalized counts can lead to

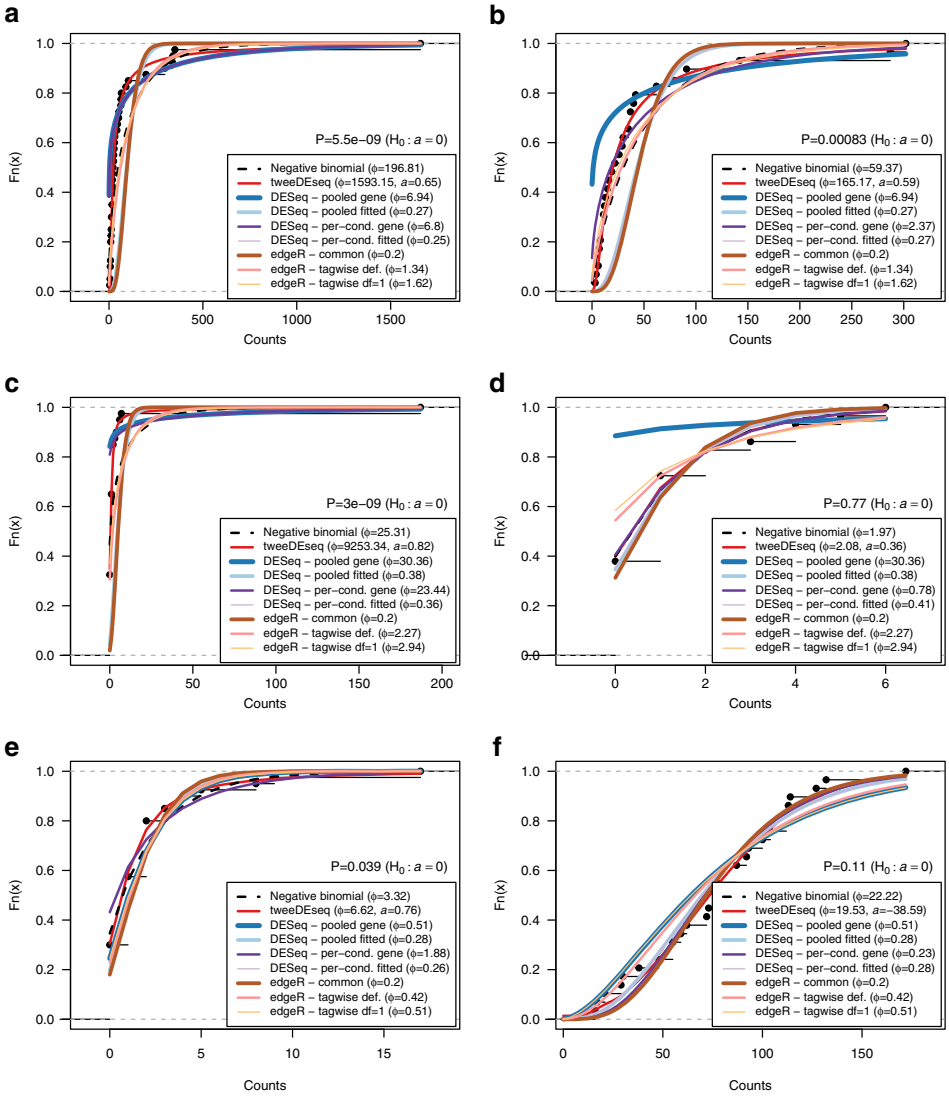

Figure 1 Fit of different count data distributions to diverse RNA-seq gene expression profiles. Fit of different count data distributions to the female **(a, c, e)** and male **(b, d, f)** RNA-seq expression profiles of genes *EEF1A2* **(a, b)**, *SCT* **(c, d)** and *NLGN4Y* **(e, f)**. All plots show the empirical cumulative distribution function (CDF) of counts (black dots) and the CDF estimated by a pure negative binomial model (black dashed line), a Poisson-Tweedie model (red line) obtained with `tweeDEseq` and several moderated negative binomial models obtained with different parameter configurations of `DESeq` and `edgeR`. Estimated dispersions, and shape in the case of `tweeDEseq`, are indicated in the legend. Above the legend, the *P*-value of the test of goodness-of-fit to a negative binomial distribution is shown. According to this test, expression profiles in panels **(a, b, c** and **e)** do not follow a negative binomial distribution. Female samples display non-negative binomial features such as a heavy-tail **(a, c)** and zero-inflation **(c, e)**. Gene *NLGN4Y* is documented in the literature as a gene with sex-specific expression, while the other two are not (*EEF1A2* is a housekeeping gene and *SCT* is an endocrine hormone peptide in chromosome 11 that controls secretions in the duodenum).

quite different MA-plots depending on the normalization method, thus potentially affecting DE detection power and accuracy (Figure 2).

The statistical methods proposed in this paper are implemented in a package for the statistical software R, called `tweeDEseq` which forms part of the Bioconductor project [13] at http://www.bioconductor.org. We have also created an experimental data package, called `tweeDEseqCountData`, which contains the previously described data set and is also available at the same URL. All results presented in the paper were obtained using these and other packages from R version 2.15.1 and Bioconductor version 2.11, and can be reproduced through the scripts available as Additional file 1 to this article.

Review of competing methods

There is currently a large body of literature on DE analysis methods for RNA-seq data [5-11,14], nearly all of them based on the NB distribution and developed to deliver their best performance with few replicates. Anders *et al.* (2010) [7] argued that for large number of individuals "... questions of data distribution could be avoided by using non-parametric methods, such as rank-based permutation tests". However, rank-based methods require similar count data distributions between sample groups. Due to the large variability across groups [15] captured by RNA-seq data, this assumption will most likely be broken in this context. For example, panels e-f in Figure 1 illustrate the case of gene *NLGN4Y* (ENSG00000165246), a gene located in the male-specific region of chromosome Y and reported to have sex-specific expression, which shows remarkably different count data distributions between male and female samples. Permutation tests are also underpowered since distribution tails are not well estimated (due to the large dynamic range), which is important when correcting for multiple testing.

In this paper we will focus our comparisons on the two most widely used methods for DE analysis of RNA-seq data, `edgeR` [6,8,10] (version 3.0.8) and `DESeq` [7] (version 1.10.1) and explore those parameter configurations in these methods that we found most relevant for large RNA-seq data sets, according to the available documentation. Both, `edgeR` and `DESEq`, assume that expression profiles from RNA-seq data follow an NB distribution and borrow information across genes to first estimate a common dispersion parameter. Then, for each gene, they estimate its genewise dispersion and moderate it towards the common one. The way in which this moderation takes place depends on the method and its configuration parameters. `DESeq` [7] allows switching between common (`sharingMode="fit-only"`) and genewise (`sharingMode="gene-est-only"`) dispersions. It provides a straightforward strategy (`sharingMode="maximum"`, default configuration) to

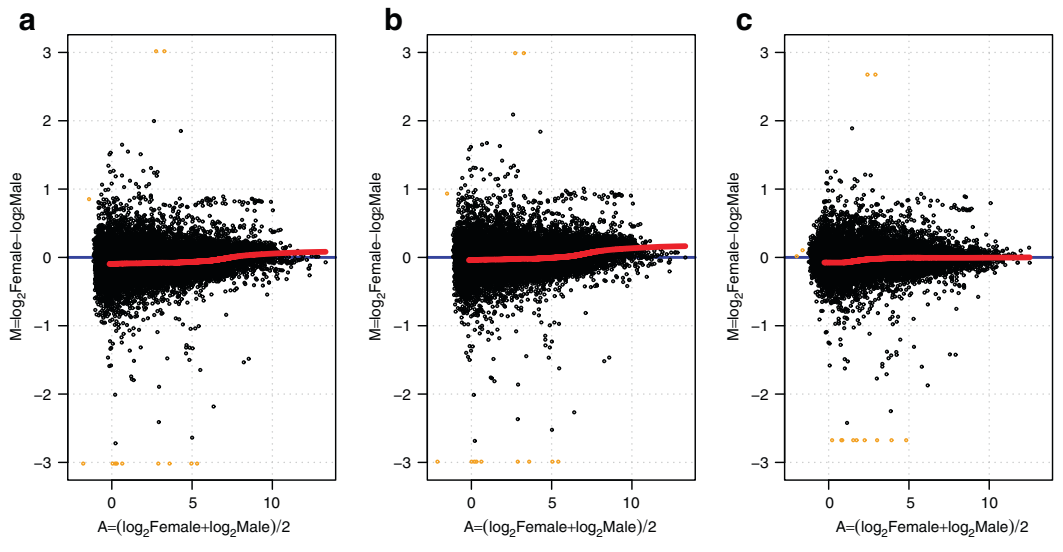

Figure 2 Count data normalization. MA-plots of the count data corresponding to the YRI samples from Pickrell [12] *et al.* (2010) after applying the following normalization methods: **(a)** raw count data without any normalization; **(b)** normalization with the `edgeR` [2] package; and **(c)** normalization with the `cqn` [4] package. The *x*-axis (A) shows the average expression throughout female and male samples in $\log_2$ scale and the *y*-axis (M) shows the magnitude of the $\log_2$-fold change between female and male samples.

choose between common and genewise dispersions by taking the largest value for each gene. `edgeR` allows one to calibrate, using the `prior.df` parameter, the transition from a purely genewise dispersion estimate (small values of `prior.df`) to the common one (large values of `prior.df`) by using an empirical Bayes approach. By default `prior.df=20` which implies that a large weight is given to the common dispersion. However, according to the documentation, if the number of samples is large, the common dispersion becomes less important in the moderation step. Additional options in `DESeq` and `edgeR` that may be relevant in the context of large RNA-seq data sets are, in the case of `DESeq`, whether dispersions are estimated from the entire pool of samples (`method="pooled"`, its default) or separately per sample group (`method="per-condition"`). In the case of `edgeR`, whether the DE test is performed using a likelihood ratio test (`glmLRT()` function) or a quasi-likelihood F-test [8] (`glmQLFTest()` function), after dispersions are estimated. Table 1 summarizes these eight combinations of methods and parameter configurations and contains the key to the terms used in some figures to distinguish among them.

Different gene expression dynamics require different distributional assumptions on count data

We assessed the goodness-of-fit of every expression profile in the LCL RNA-seq data to an NB distribution (see Methods) by means of quantile-quantile (Q-Q) plots (Figure 3) and about 10% of the genes show a substantial discrepancy with respect to the NB distribution in the counts (see right y-axis in Figure 3). Such a discrepancy is absent from data simulated from NB distributions with a similar number of genes including a small fraction of them changing between two conditions (Additional file 2: Figure S1).

This result suggests that NB distributions may be too restrictive for an important fraction of expression profiles in large RNA-seq data sets. Among the possible causes underlying the lack of fit of those genes to an NB distribution, a clear candidate is that the presence of many samples can potentially introduce features such as zero-inflation or heavy-tails (see Figure 1). So far, extensive biological replication in RNA-seq experiments has been the exception rather than the rule. However, it is becoming increasingly clear [15] that in the coming years larger RNA-seq data sets will be required to justify scientific conclusions and provide reproducible results. Therefore, we can expect to see more often gene expression profiles with emerging features, such as zero-inflation and heavy tails, that challenge RNA-seq methods based on the NB distribution.

We propose to address this problem by adopting the Poisson-Tweedie (PT) family of distributions [16] to model RNA-seq count data directly. PT distributions are described by a mean (μ), a dispersion (ϕ) and a shape (a) parameter (see Methods) and include Poisson and NB distributions, among others, as particular cases [16]. An important feature of this family is that, while the NB distribution only allows a quadratic mean-variance relationship, the PT distributions generalizes this relationship to any order [17]. We have implemented a maximum likelihood procedure for the estimation and simulation of these parameters from count data. These procedures are available in the `tweeDEseq` package through the functions `mlePoissonTweedie()`, `dPT()` and `rPT()`.

Figure 1 illustrates the flexibility of the PT distribution to accurately fit different gene expression profiles obtained from the un-normalized LCL RNA-seq data set. Left and right panels correspond to female and male samples, respectively and each row corresponds to a different gene: *EEF1A2* (ENSG00000101210), *SCT* (ENSG00000070031) and *NLGN4Y* (ENSG00000165246), respectively. Among these three genes, only *NLGN4Y* has been reported in the literature to have sex specific expresssion, while the other two are likely to lack such property since *EEF1A2* is a housekeeping gene and *SCT* is an endocrine hormone peptide in chromosome 11 that controls secretions in the duodenum. Each plot shows the empirical cumulative distribution of observed counts as

Table 1 Methods and parameter configurations compared in this paper

Key	Software	Configuration parameters
DESeqPgO	`DESEq`	`method="pooled"`, `sharingMode="per-condition"`
DESeqPmax	`DESEq`	`method="pooled"`, `sharingMode="maximum"`
DESeqCgO	`DESEq`	`method="per-condition"`, `sharingMode="per-condition"`
DESeqCmax	`DESEq`	`method="per-condition"`, `sharingMode="maximum"`
edgeRdf20	`edgeR`	common/trended/tagwise moderation regime with `prior.df=20` (default)
edgeRdf1	`edgeR`	common/trended/tagwise moderation regime with `prior.df=1`
edgeRqlfDf20	`edgeR`	common/trended/tagwise moderation regime with `prior.df=20` (default) and quasi-likelihood F-tests
edgeRqlfDf1	`edgeR`	common/trended/tagwise moderation regime with `prior.df=1` and quasi-likelihood F-tests

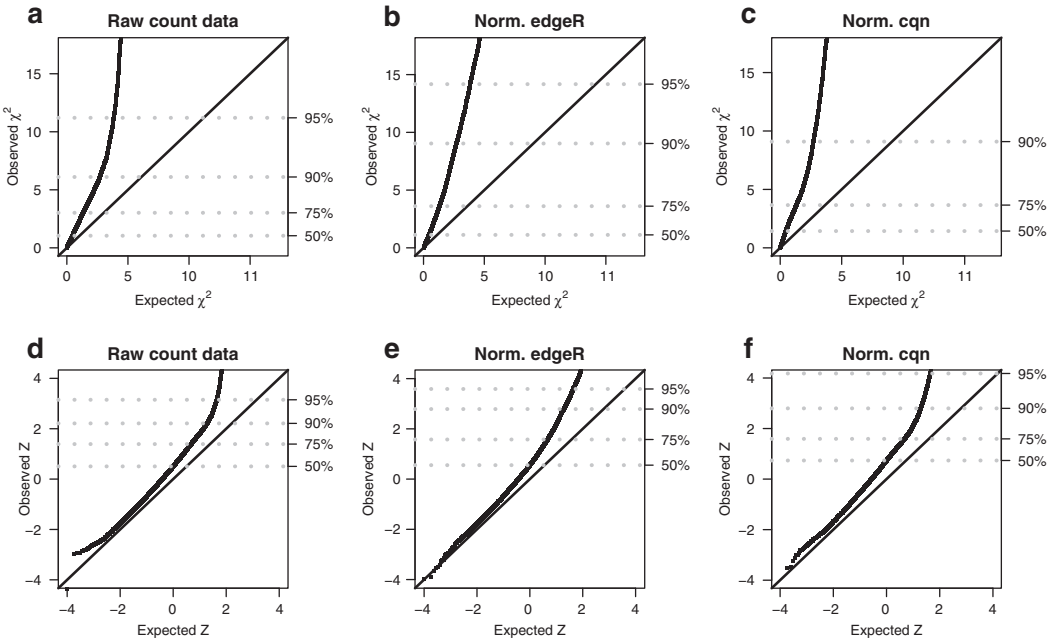

Figure 3 Goodness of fit to the negative-binomial distribution. Quantile-quantile (Q-Q) plots of the goodness-of-fit of RNA-seq expression profiles from Pickrell [12] *et al.* (2010) to a negative-binomial (NB) distribution. The right *y*-axis indicates the quantile of the observed distribution. Columns correspond to different normalization methods where **(a, d)** correspond to raw un-normalized counts, **(b, e)** normalization with `edgeR` and **(c, f)** normalization with `cqn`. The top row **(a, b, c)** contains the Q-Q plots of the χ^2 goodness-of-fit statistic while the bottom row (d, e, f) contains the same Q-Q plot mapped to a normalized Z-statistic to improve the visibility of the left tail of the distribution. Independently on how count data are normalized, about 10% of the expression profiles show a substantial discrepancy to the NB distribution.

well as the parametric cumulative distributions obtained through the estimation of parameters of the methods compared in this paper under different configurations. Note that the estimated dispersion parameter ϕ is identical between the two sample groups for `edgeR` and `DESeq` (pooled) as these approaches estimate ϕ irrespective from the sample groups. The *P*-value for testing whether the data follow an NB distribution ($H_0 : a = 0$), indicated above the legend, reveals that in several sample groups (panels a-c, e) this hypothesis is rejected ($P < 0.05$). In those cases, methods based on the NB distribution produce dispersion parameters that do not fit the data as accurately as the PT distribution. More concretely, heavy-tails present in panels a,c severely hamper the estimation of the pure NB and the common dispersion. These can be improved using a parameter configuration more suited to large sample sizes. However, this results in a poor estimate of zero-inflation in panels c-e.

The main difference between the PT and NB distributions lies in the additional "shape" parameter a of the PT distribution which provides further flexibility (see Methods). Using the LCL data processed with different normalization methods, we show in Figure 4 all values of the shape parameter a for every gene as function of its mean expression level, illustrating the huge variability of this parameter in RNA-seq count data. This wide range of values involves distinct possible distributional assumptions [16] beyond Poisson and NB, such as Poisson-Inverse Gaussian, Pólya-Aeppli or Neyman type A. Similarly to the MA-plots of Figure 2, the `cqn` normalization method seems to make the largest impact on count data and, in this case, on the shape parameter.

We have investigated whether this diversity of count distributions underlying RNA-seq data is related to different expression dynamics in genes. Using the test for the goodness of fit to an NB distribution (see Methods) we have considered as NB those genes that do not reject the null hypothesis at $P > 0.2$ and as clear-cut non-NB genes those with $P < 2^{-16}$. By mapping all these genes to the Gene Expression Barcode catalog [18] (see Methods) we obtained an independent and unbiased estimation of their expression breadth. The results in Figure 5 suggest that the expression breadth of non-

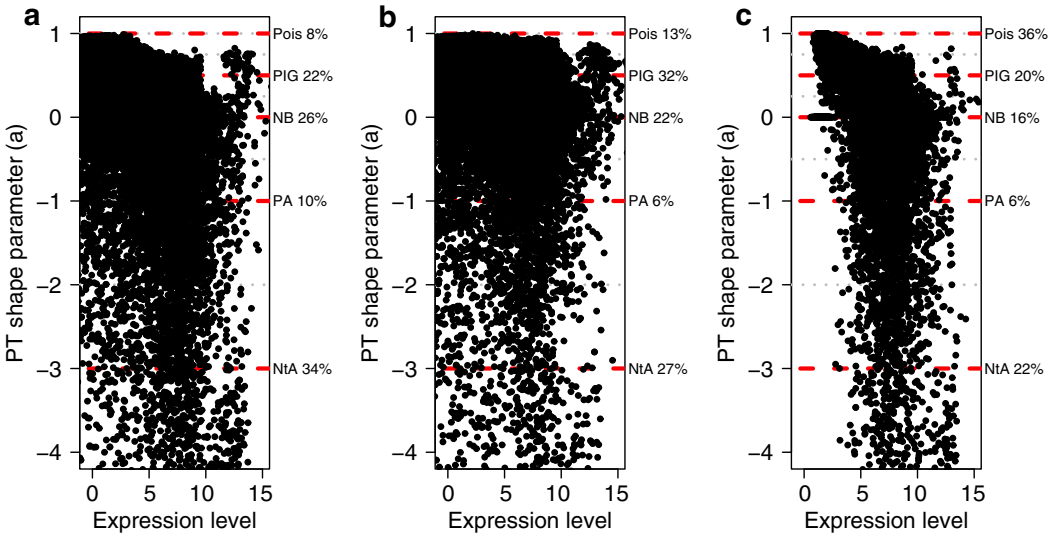

Figure 4 Distribution of the Poisson-Tweedie shape parameter as function of the mean expression level. Estimated Poisson-Tweedie shape parameter *a* as function of the mean expression level for each gene. Red dashed lines indicate the value of *a* corresponding to each specific distribution within the Poisson-Tweedie family, denoted by Pois (Poisson), PIG (Poisson-Inverse Gaussian), NB (negative binomial), PA (Pólya-Aeppli) and NtA (Neyman type A). The right *y*-axis indicates the percentage of genes around specific *a* values bounded by dotted grey lines. Data from Pickrell [12] *et al.* (2010) are shown without any normalization **(a)**, normalized with `edgeR` [2] **(b)**, and normalized with `cqn` [4] **(c)**.

NB genes approaches that of housekeeping genes closer than NB genes do, irrespective of the normalization method.

In fact, Fisher's exact tests for enrichment of non-NB genes among human housekeeping genes are significant ($P < 1.24^{-6}$) for every normalization method (see Additional file 2: Table S1). These observations suggest that genes with different expression dynamics can produce different count data distributions, and underscore the flexibility of the PT statistical model to capture

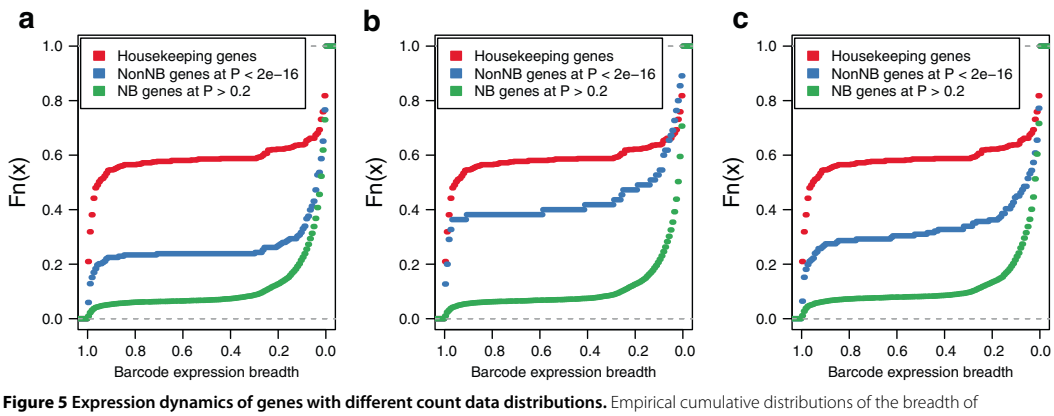

Figure 5 Expression dynamics of genes with different count data distributions. Empirical cumulative distributions of the breadth of expression estimated through the Barcode [18] database, for genes that do not reject the null hypothesis of a negative-binomial (NB) distribution in a test for the goodness of fit at $P > 0.2$ (green lines), genes that do reject such a null hypothesis at $P < 2^{-16}$ (blue lines) and housekeeping genes retrieved from literature [19] (red lines). Data from Pickrell [12] *et al.* (2010) are shown without any normalization **(a)**, normalized with `edgeR` [2] **(b)** and normalized with `cqn` [4] **(c)**. These plots show that, independently of the normalization method, non-NB genes at such significance level of discrepancy with respect to the NB distribution approach closer the expression dynamics of housekeeping genes than genes with expression profiles following the NB distribution.

these dynamics revealed by extensively-replicated RNA-seq experiments.

Accurately testing differential expression

For the purpose of a DE analysis between two groups of samples, we have developed a two-sample PT-test for differences in means (see Methods) implemented through the function `tweeDE()` in the `tweeDEseq` package. We will assess the accuracy of this PT-based test using the LCL data as well as synthetic count data from two different simulation studies. The first simulation study with synthetic data provides an assessment of the type-I error rate under four different scenarios involving distinct count data distributions between sample groups (see Additional file 2: Table S2 for a description of them). Here we compare `tweeDEseq` with the configurations of `edgeR` and `DESeq` which are closer to a straightforward NB model. Additional file 2: Figures S2 to S5 show that `tweeDEseq` properly controls the nominal probability of a type-I error while `edgeR`, `DESeq` and non-parametric tests (U Mann-Withney and permutation) fail to do so when data are not simulated from NB distributions. As expected, these methods perform correctly when data are generated under an NB model (see Additional file 2: Figure S5) as expected. Additional file 2: Figure S6 also shows that in the calculation of very low *P*-values, `tweeDEseq` clearly outperforms permutations tests. In order to provide a practical recommendation on the minimum sample size required by `tweeDEseq` to yield accurate results we have estimated the probability of a type-I error across different sample sizes. Additional file 2: Figure S7 indicates that 15 samples per group should be sufficient for `tweeDEseq` to correctly control for a nominal significance level $\alpha = 0.05$.

In the second simulation study we have first assessed the accuracy of the *P*-value distribution under the null hypothesis of no differential expression with real RNA-seq data by making repeatedly two-sample group comparisons within males and within females samples such that we recreate the null hypothesis of no DE with real RNA-seq data and no DE gene should be expected to be found. The raw *P*-value distributions from such analysis should ideally be uniform.

We have formally tested this hypothesis for every gene by means of a Kolmogorov-Smirnov (KS) goodness-of-fit test to a uniform distribution and examine the resulting *P*-value distribution by means of Q-Q plots displayed in Figure 6. Under the null hypothesis that all genes are not DE, the KS *P*-values should lie along the diagonal of the Q-Q plot. The figure, however, shows large discrepancies to this criterion by some of the methods and configuration parameters, indicating that they may not be adequate for large RNA-seq data sets.

The method introduced in this paper, `tweeDEseq`, is consistently closer to the diagonal than every other method throughout the two male and female comparisons and the two normalization methods. More informally, the visual inspection of the histograms of raw *P*-values given in Additional file 2: Figure S8 also reveals that `tweeDEseq` provides *P*-value distributions closer to the uniform under the null hypothesis of no DE simulated from extensively replicated real RNA-seq data.

As other authors have shown, in the context of analysis of RNA-seq data with very limited sample size [8], small deviations from uniformity of *P*-values under the null hypothesis can substantially affect FDR estimates of DE genes. We have also assessed the calibration of *P*-values and false discovery rates (FDR) with synthetic count data of similar dimensions to the RNA-seq LCL data set, concretely with $p = 20,000$ genes and $n = 70$ samples. Working with this type of data allows to assess FDR estimates for a known subset of DE genes under a variety of simulated scenarios, which we defined by considering the combination of three different amounts of DE genes (100, 1000 and 2000) and three different magnitudes of fold-change (1.5, 2 and 4-fold). Similarly to [8], data were simulated from a hierarchical gamma-Poisson model with and without simulated library factors (see Methods).

From every simulated data set, raw *P*-values for the two-sample DE test were obtained with each method and configuration parameters. Using the `qvalue` Bioconductor package [21] we estimated *q*-values and the fraction of DE genes from each *P*-value distribution. Q-values provide a nominal estimation of the FDR for each gene and in Figures 7 and 8 we show the empirical FDR (eFDR) as a function of the nominal *q*-values for the simulations with constant and variable library factors, respectively. The dashed diagonal line indicates a correct calibration of *P*-values whose nominal FDR equals the observed eFDR. Lines above the diagonal correspond to liberal DE analysis methodologies and below to conservative ones.

To facilitate the comparison of methods across all simulated data sets we have calculated the mean squared error (MSE) between the eFDR and the nominal FDR and ranked the methods by increasing MSE. In Tables 2 and 3 we can find the MSE values and in Tables 4 and 5 the corresponding ranks of the methods according to the MSE values. As it follows from the rankings in Tables 4 and 5, `tweeDEseq` provides the best calibrated *P*-values in most of the simulated data sets.

The previous calculations of *q*-values with the `qvalue` package [21] provide us also with estimates $\hat{\pi}_0$ of the fraction of genes under the null hypothesis of no differential expression. This, in turn, allows one to derive an

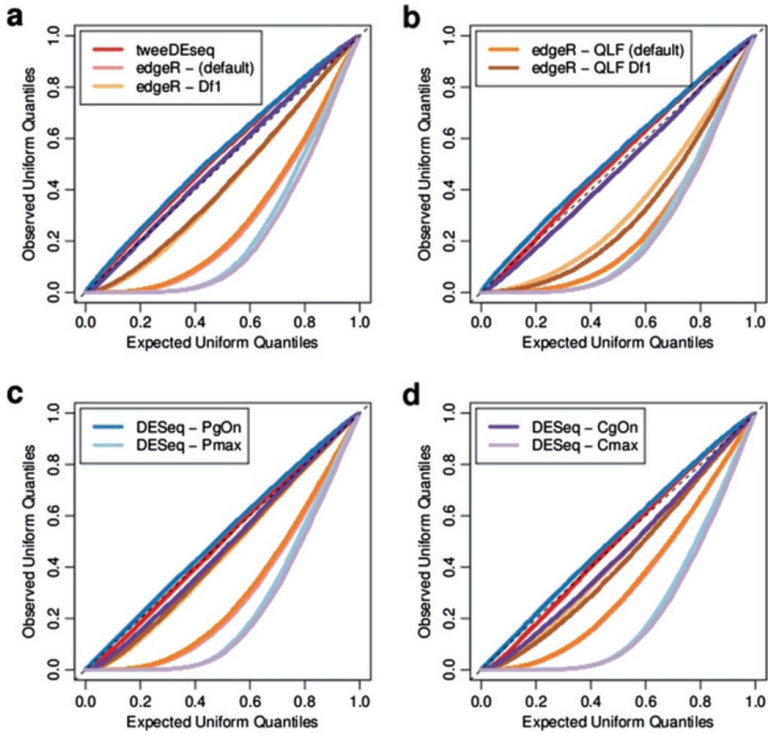

Figure 6 Quantile-quantile (Q-Q) plots for the goodness-of-fit of null-hypothesis *P*-values to an uniform distribution. Using the results displayed in Additional file 2: Figure S8 and performing as described by Leek *et al.* (2007) [20], for each gene, the distribution of *P*-values throughout the 100 simulations was tested for its goodness-of-fit to an uniform distribution using a Kolmogorov-Smirnov test. Q-Q plots in this figure show for all genes the resulting *P*-values of the previous test which, under the null hypothesis of no differential expression, should be uniformly distributed too and lead to lines lying on the diagonal. Panels **a-b** show results from female vs female comparisons and **c-d** from male vs male comparisons, while a,c correspond to un-normalized data and b,d to data normalized with the `cqn` [4]. The method introduced in this paper, `tweeDEseq`, is on average closer to the diagonal throughout the four simulations, closely followed by `DESeq` when `sharingMode="gene-est-only"` and either `method="per-condition"` or `method="pooled"`.

estimated number of DE genes as $p(1 - \hat{\pi}_0)$ with p being the total number of genes. In principle, more precise *P*-values both under the null and the alternative hypotheses should provide more accurate estimates of the number of DE genes. We show such an assessment for the previous simulations in Additional file 2: Figures S9 and S10. To summarize those results we have divided each estimate of the number of DE genes by their actual simulated number of DE genes and aggregate those ratios throughout the different simulation scenarios to ease the comparison among the methods. We find this comparison in Figure 9 and it follows that `tweeDEseq` produces *P*-values that lead to the most accurate estimation of the number of DE genes, closely followed by `edgeR` with `prior.df=1` when library factors are not held constant. In both settings, `DESeq` leads to extremely conservative estimates of the number of DE genes.

Identification of sex-specific gene expression in lymphoblastoid cell lines

Assessing performance of DE analysis methods without using simulated data is a challenging problem due to the difficulty of knowing or ensuring the exact differential concentration of RNA molecules in the analysed samples. In this respect, sex-specific expression constitutes a useful system to assess the accuracy of DE detection methods due to the vast literature on genes contributing to gender-specific traits. For this reason, in order to illustrate the accuracy of `tweeDEseq` with real RNA-seq data, we have searched for genes changing significantly their expression between female and male individuals of the RNA-seq experiments on LCLs analyzed in this paper. Again, we have compared different normalization procedures and parameter configurations of `edgeR` and `DESeq`. Next to considering the raw un-normalized data and the data

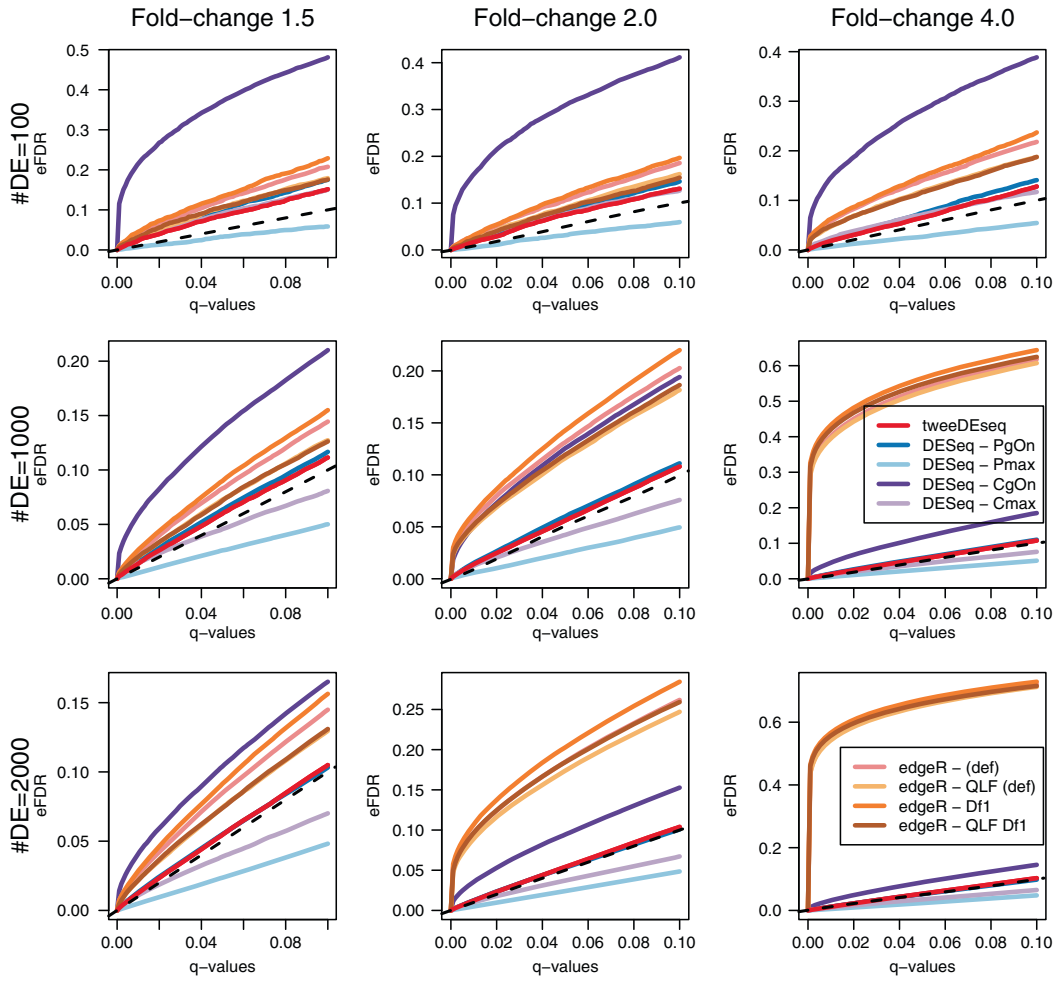

Figure 7 Empirical FDR values for simulated data with constant library factors. Empirical FDR values on the *y*-axis as function of nominal *q*-values on the *x*-axis calculated from data simulated with $p = 20,000$ genes, $n = 70$ samples and constant library factors. Each row and column corresponds, respectively, to a different number of DE genes and magnitude of the fold-change. The method introduced in this paper, `tweeDEseq`, is consistently closer to the diagonal than other methods throughout the different simulations.

normalized with `cqn`, TMM normalization was used for `edgeR` and `tweeDEseq`, while `DESeq` was used with its own normalization method. We have used a single significance cutoff of FDR < 0.1 at which genes were called DE. Since LCLs come from a non-sexually dimorphic tissue and are outside their original biological context, the fraction of sex-specific expression changes we could expect should be rather small.

In an attempt to verify the accuracy of these lists of DE genes between female and male individuals, we searched for genes reported in the literature to be potential contributors to sexually dimorphic traits. This list of genes with documented sex-specific expression was obtained from genes in chromosome X that escape X-inactivation [22] and from genes in the male-specific region of the Y chromosome [23] (see Methods). This resulted in a gold-standard set of 95 genes mapping to Ensembl Gene Identifiers (release 63), which we shall denote by XiE and MSY genes, depending on their origin. For every predicted set of DE genes by each combination of DE detection method and normalized data set, we calculated precision and recall with respect to the gold-standard, and the *F*-measure which summarizes the trade-off between these two diagnostics.

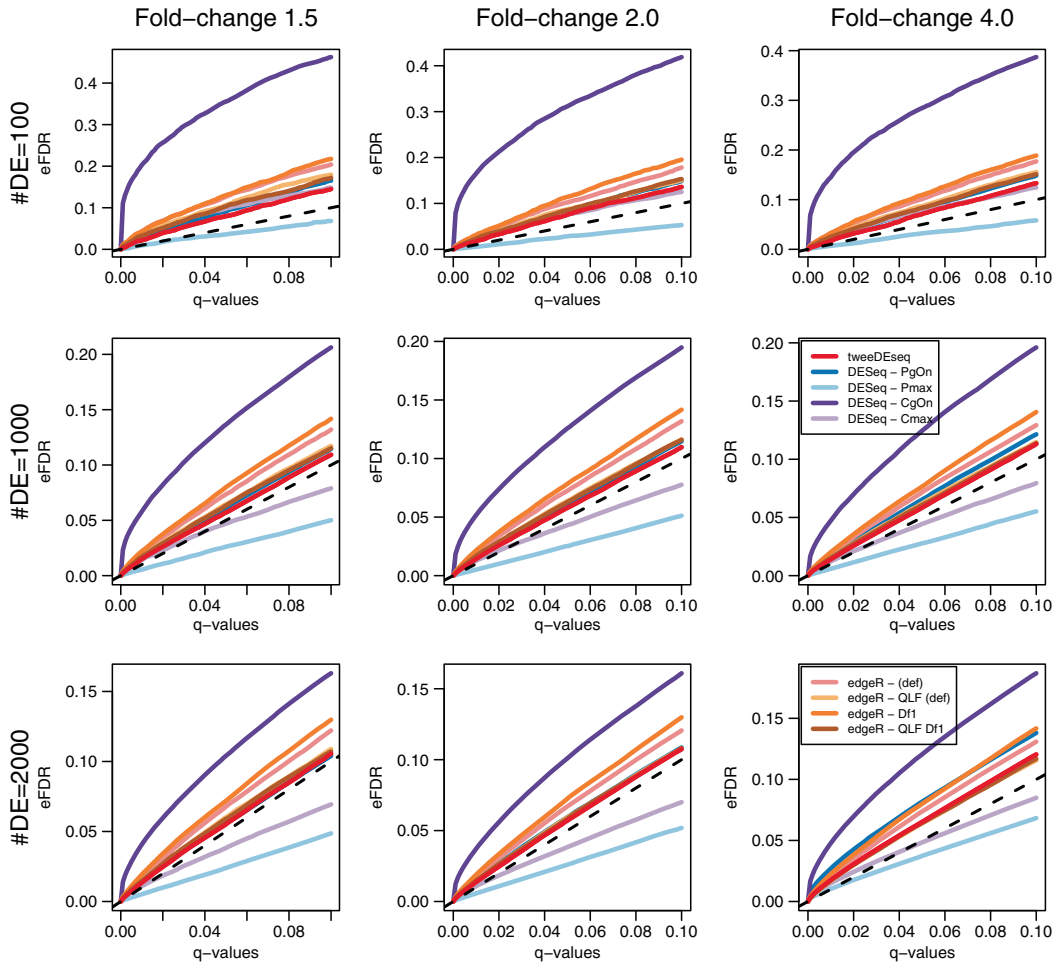

Figure 8 Empirical FDR values for simulated data with variable library factors. Empirical FDR values on the *y*-axis as function of nominal *q*-values on the *x*-axis calculated from data simulated with $p = 20,000$ genes, $n = 70$ samples and variable library factors. Each row and column corresponds, respectively, to a different number of DE genes and magnitude of the fold-change. The method introduced in this paper, `tweeDEseq`, is consistently closer to the diagonal than other methods throughout the different simulations.

In Figure 10 we can see that `tweeDEseq` provides better performance than the other competing methods under different parameter configurations. The improvement is small with respect to the second best-performing method and parameter configuration but we would like to emphasize that `tweeDEseq` does not require any informed decision on a parameter configuration, as opposed to `edgeR` and `DESeq`. To assess the robustness of this figure, we have run this comparative assessment with a more stringent filter on lowly expressed genes and, as Additional file 2: Figure S11 shows, `tweeDEseq` keeps performing better than the other methods, this time however only when data are normalized.

In Additional file 2: Table S3 we report the complete list of 55 DE genes detected by `tweeDEseq` from the data normalized with `cqn`, which is when it yields the best precision-recall tradeoff. More than a half of genes in this list (32) are located in either the X or Y chromosomes and where the first 10 with largest fold-change contain 7 from the gold-standard set of MSY and XiE genes. Among the other 3, we find *TTTY15*, a testis-specific non-coding transcript from the Y chromosome and the other two lack functional annotation in Ensembl release 63.

Table 2 Mean squared error of false discovery rates under constant library factors

#DE	Rank	1.5-fold change		2-fold change		4-fold change	
		MSE	Method	MSE	Method	MSE	Method
100	1	0.050	DESeq - Pmax	0.050	DESeq - Pmax	0.031	tweeDEseq
	2	0.122	tweeDEseq	0.053	tweeDEseq	0.037	DESeq - Cmax
	3	0.155	DESeq - Cmax	0.065	DESeq - Cmax	0.070	DESeq - Pmax
	4	0.257	DESeq - PgOn	0.104	DESeq - PgOn	0.072	DESeq - PgOn
	5	0.306	edgeR - QLF Df1	0.138	edgeR - QLF Df1	0.430	edgeR - QLF Df1
	6	0.313	edgeR - QLF (def)	0.177	edgeR - QLF (def)	0.452	edgeR - QLF (def)
	7	0.558	edgeR - (def)	0.336	edgeR - (def)	0.790	edgeR - (def)
	8	0.755	edgeR - Df1	0.431	edgeR - Df1	0.957	edgeR - Df1
	9	9.688	DESeq - CgOn	6.232	DESeq - CgOn	5.133	DESeq - CgOn
1000	1	0.008	tweeDEseq	0.004	tweeDEseq	0.004	tweeDEseq
	2	0.008	DESeq - Cmax	0.008	DESeq - PgOn	0.007	DESeq - PgOn
	3	0.016	DESeq - PgOn	0.015	DESeq - Cmax	0.014	DESeq - Cmax
	4	0.043	edgeR - QLF Df1	0.087	DESeq - Pmax	0.081	DESeq - Pmax
	5	0.045	edgeR - QLF (def)	0.413	edgeR - QLF (def)	0.429	DESeq - CgOn
	6	0.082	DESeq - Pmax	0.459	edgeR - QLF Df1	21.358	edgeR - QLF (def)
	7	0.105	edgeR - (def)	0.532	DESeq - CgOn	22.208	edgeR - (def)
	8	0.155	edgeR - Df1	0.639	edgeR - (def)	23.401	edgeR - QLF Df1
	9	0.735	DESeq - CgOn	0.835	edgeR - Df1	25.004	edgeR - Df1
2000	1	0.002	DESeq - PgOn	0.001	DESeq - PgOn	0.000	DESeq - PgOn
	2	0.002	tweeDEseq	0.001	tweeDEseq	0.001	tweeDEseq
	3	0.025	DESeq - Cmax	0.031	DESeq - Cmax	0.036	DESeq - Cmax
	4	0.053	edgeR - QLF (def)	0.090	DESeq - Pmax	0.093	DESeq - Pmax
	5	0.056	edgeR - QLF Df1	0.183	DESeq - CgOn	0.140	DESeq - CgOn
	6	0.093	DESeq - Pmax	1.444	edgeR - QLF (def)	34.551	edgeR - QLF (def)
	7	0.113	edgeR - (def)	1.702	edgeR - QLF Df1	35.365	edgeR - (def)
	8	0.169	edgeR - Df1	1.724	edgeR - (def)	35.468	edgeR - QLF Df1
	9	0.271	DESeq - CgOn	2.219	edgeR - Df1	36.929	edgeR - Df1

Data in this table correspond to the mean squared error (MSE) values between the empirical false discovery rates (eFDR) and the nominal *q*-values obtained from the simulation study shown in Figure 7 in which library factors were held constant.

Reproducibility with respect to microarray data

The YRI LCL samples we have analyzed have been previously assayed using microarray chips [24] and this enables a comparison between the gene expression read out of both technologies. In particular, we wanted to assess the degree of reproducibility of the significance levels of DE. While there may be many aspects from both technologies that can potentially bound the extent to which we can reproduce rankings of DE, we postulate that more accurate *P*-values in DE genes should lead to higher reproducibility of significance levels of DE genes.

With this purpose, we applied `limma` [25] on the microarray data and called genes DE at 10% FDR, just as we did with RNA-seq data, and then compared the $-\log_{10}$ units of the raw *P*-values from DE genes called in RNA-seq by each DE detection method to the $-\log_{10}$ *P*-value units from genes called DE by `limma`. In Additional file 2: Figure S12 we show this comparison for every gene that is called DE either by `limma` in microarray data or by the other compared method in RNA-seq data. Although we can observe a significant linear relationship between *P*-values in every compared method, the low fraction of variability explained by the fitted linear model ($R^2 < 0.25$) in every of those comparisons indicates a rather poor level of reproducibility for every method.

A closer look to genes in that figure indicates that the lack of reproducibility mostly comes from genes called DE by one method and technology but not by the other (dots close to zero in either the x or the y-axis). There may be many reasons, unrelated to the DE detection method itself, why a gene is not called simultaneously DE in two completely independent RNA-seq and microarray

Table 3 Mean squared error of false discovery rates under variable library factors

#DE	Rank	1.5-fold change		2-fold change		4-fold change	
		MSE	Method	MSE	Method	MSE	Method
100	1	0.030	DESeq - Pmax	0.059	DESeq - Cmax	0.046	tweeDEseq
	2	0.099	tweeDEseq	0.064	tweeDEseq	0.055	DESeq - Pmax
	3	0.189	DESeq - Cmax	0.072	DESeq - Pmax	0.057	DESeq - Cmax
	4	0.194	DESeq - PgOn	0.116	DESeq - PgOn	0.106	DESeq - PgOn
	5	0.258	edgeR - QLF Df1	0.129	edgeR - QLF Df1	0.124	edgeR - QLF Df1
	6	0.348	edgeR - QLF (def)	0.129	edgeR - QLF (def)	0.153	edgeR - QLF (def)
	7	0.581	edgeR - (def)	0.273	edgeR - (def)	0.290	edgeR - (def)
	8	0.667	edgeR - Df1	0.420	edgeR - Df1	0.380	edgeR - Df1
	9	8.882	DESeq - CgOn	6.429	DESeq - CgOn	5.217	DESeq - CgOn
1000	1	0.005	tweeDEseq	0.006	tweeDEseq	0.008	tweeDEseq
	2	0.009	DESeq - Cmax	0.012	DESeq - Cmax	0.010	DESeq - Cmax
	3	0.013	DESeq - PgOn	0.012	DESeq - PgOn	0.011	edgeR - QLF Df1
	4	0.016	edgeR - QLF Df1	0.016	edgeR - QLF Df1	0.013	edgeR - QLF (def)
	5	0.019	edgeR - QLF (def)	0.017	edgeR - QLF (def)	0.024	DESeq - PgOn
	6	0.054	edgeR - (def)	0.051	edgeR - (def)	0.045	edgeR -(def)
	7	0.083	DESeq - Pmax	0.082	DESeq - Pmax	0.067	DESeq - Pmax
	8	0.087	edgeR - Df1	0.083	edgeR - Df1	0.077	edgeR - Df1
	9	0.700	DESeq - CgOn	0.545	DESeq - CgOn	0.529	DESeq - CgOn
2000	1	0.003	tweeDEseq	0.004	tweeDEseq	0.005	DESeq - Cmax
	2	0.003	DESeq - PgOn	0.006	edgeR - QLF Df1	0.017	edgeR - QLF (def)
	3	0.006	edgeR - QLF Df1	0.006	edgeR - QLF (def)	0.018	edgeR - QLF Df1
	4	0.007	edgeR - QLF (def)	0.007	DESeq - PgOn	0.023	tweeDEseq
	5	0.025	DESeq - Cmax	0.024	DESeq - Cmax	0.029	DESeq - Pmax
	6	0.028	edgeR - (def)	0.026	edgeR - (def)	0.053	edgeR - (def)
	7	0.049	edgeR - Df1	0.047	edgeR - Df1	0.088	edgeR - Df1
	8	0.091	DESeq - Pmax	0.077	DESeq - Pmax	0.092	DESeq - PgOn
	9	0.267	DESeq - CgOn	0.238	DESeq - CgOn	0.465	DESeq - CgOn

Data in this table correspond to the mean squared error (MSE) values between the empirical false discovery rates (eFDR) and the nominal *q*-values obtained from the simulation study shown in Figure 8 in which library factors were variable.

Table 4 Rankings of methods by the mean squared error of false discovery rates under constant library factors

Method	#DE = 100			#DE = 1000			#DE = 2000		
	1.5 FC	2 FC	4 FC	1.5 FC	2 FC	4 FC	1.5 FC	2 FC	4 FC
tweeDEseq	2	2	1	1	1	1	2	2	2
DESeq - PgOn	4	4	4	3	2	2	1	1	1
DESeq - Pmax	1	1	3	6	4	4	6	4	4
DESeq - CgOn	9	9	9	9	7	5	9	5	5
DESeq - Cmax	3	3	2	2	3	3	3	3	3
edgeR - (def)	7	7	7	7	8	7	7	8	7
edgeR - QLF (def)	6	6	6	5	5	6	4	6	6
edgeR - Df1	8	8	8	8	9	9	8	9	9
edgeR - QLF Df1	5	5	5	4	6	8	5	7	8

Data in this table correspond to the rankings of every method by the mean squared error (MSE) values shown in Table 1.

Table 5 Rankings of methods by the mean squared error of false discovery rates under variable library factors

Method	#DE = 100			#DE = 1000			#DE = 2000		
	1.5 FC	2 FC	4 FC	1.5 FC	2 FC	4 FC	1.5 FC	2 FC	4 FC
tweeDEseq	2	2	1	1	1	1	1	1	4
DESeq - PgOn	4	4	4	3	3	5	2	4	8
DESeq - Pmax	1	3	2	7	7	7	8	8	5
DESeq - CgOn	9	9	9	9	9	9	9	9	9
DESeq - Cmax	3	1	3	2	2	2	5	5	1
edgeR - (def)	7	7	7	6	6	6	6	6	6
edgeR - QLF (def)	6	6	6	5	5	4	4	3	2
edgeR - Df1	8	8	8	8	8	8	7	7	7
edgeR - QLF Df1	5	5	5	4	4	3	3	2	3

Data in this table correspond to the rankings of every method by the mean squared error (MSE) values shown in Table 2.

experiments on the same biological material, such as different isoforms being probed in the microarray and summarized in RNA-seq or differences in sample preparation. Therefore, for our current goal of assessing reproducibility of DE detection methods, we believe it makes sense to restrict this comparison to those genes that are called DE by both, `limma` in microrray data and the corresponding method in RNA-seq data.

We can find this restricted comparison in Figure 11 which reveals that in this case only `tweeDEseq` attains a significant ($P < 0.05$) linear fit with respect to the P-values from `limma` with a level of reproduciblity ($R^2 = 0.6$) substantially larger (46% increase) than the second best method (`DESeq - PgO`) with $R^2 = 0.41$.

Finally, we have carried out a comparison between the entire output of DE genes obtained with `tweeDEseq` in RNA-seq data with the entire output DE genes obtained with `limma` in microarray data. In Figure 12 we show the resulting volcano plots where we have highlighted with black dots those genes that are exclusively profiled by each technology. As the figure suggests, many more of these genes occur in RNA-seq than in microrray, one remarkable case being the *XIST* gene which shows the largest fold-change and significance level and corresponds

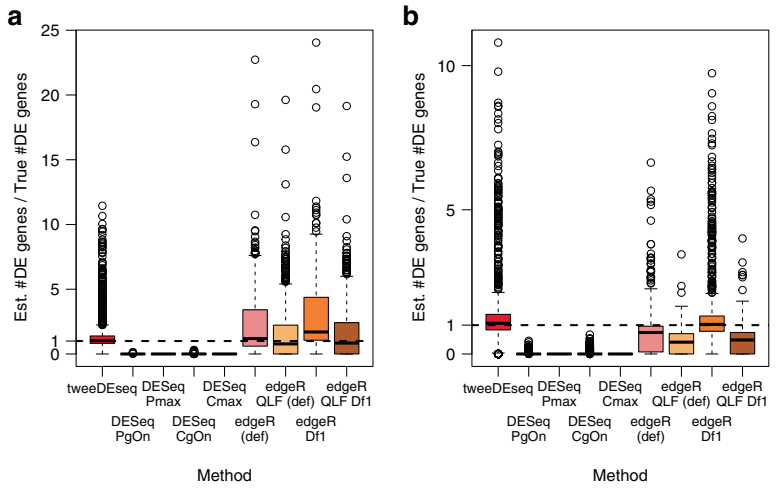

Figure 9 Estimation of the number of differentially expressed (DE) genes from simulated data. Boxplots of ratios of estimated to true numbers of DE genes obtained from data simulated from a hierarchical gamma-Poisson model with constant **(a)** and variable **(b)** library factors. This figure summarizes the results in Additional file 2: Figures S9 and S10 reporting estimated numbers of DE genes under different simulated scenarios of number or true DE genes and fold-change. The horizontal dash line at ratio one corresponds to the best performance where the estimated number of DE genes matches the true number.

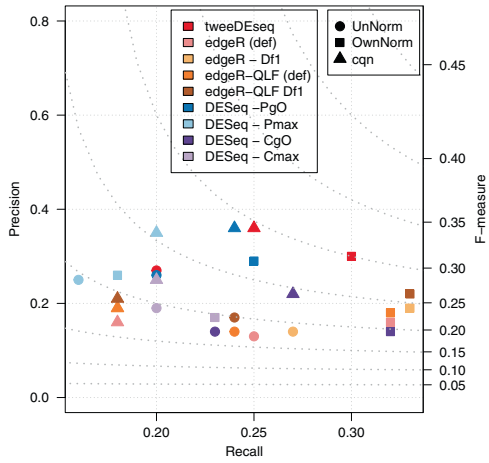

Figure 10 Precision and recall comparison on the LCL RNA-seq data. Precision (*y*-axis) and recall (*x*-axis) values for genes called DE at 1% FDR by different DE detection methods and configuration parameters. The right *y*-axis indicates values of the *F*-measure shown by dot lines. As the figure shows, `tweeDEseq` provides higher *F*-measure values than other methods indicating a better precision-recall tradeoff.

to the X-inactive specific non-coding RNA gene which acts as one of the key regulators in silencing one of the copies of chromosome X in females. Blue and red circles denote MSY and XiE genes, respectively. As expected, all MSY and XiE DE genes report significantly higher expression in males and females, respectively, except for the XiE gene *NLGN4X* in RNA-seq, likely due to low expression from the inactive X chromosome in female samples [26]. Surprisingly the volcano plots show that `limma` on this microarray data set is able to detect a few more such genes than `tweeDEseq` on RNA-seq data. Last, but not least, an important difference between the volcano plots of Figure 12 is the fact that expression changes larger than 2-fold in these microarray data are nearly synonymous of statistical significance while with RNA-seq a sizeable fraction of genes with 2-fold or larger changes show very poor significance levels. This is likely due to the larger variability of gene expression measurements in RNA-seq experiments with many samples and underscores the importance of using methods that properly assess the statistical significance of the observed changes.

Conclusions

The increased amount of biological variability revealed by extensive replication in RNA-seq experiments brings new challenges to the task of identifying genes whose change in expression is both, biologically and statistically significant. In microarray data, large fold-changes derived from large data sets were nearly synonymous of statistical significance. The volcano plots in Figure 12 and the examples of specific genes in Figure 1 illustrate why this is not true anymore with RNA-seq count data. Those figures unveil that one of these new challenges is to distinguish statistically significant changes among those that are already large in magnitude. In this paper we provide an approach to this problem by using the PT family of distributions, showing that it captures a much richer diversity of expression dynamics in RNA-seq count data than the statistical models based in the NB distributions alone (see Figures 4 and 5). We have implemented a two-sample PT-test in a software package for R, called `tweeDEseq`, for detecting DE genes and demonstrated with simulations that produces more accurate *P*-value distributions that lead to better calibrated *q*-values and FDR estimates.

We have made an attempt to assess DE detection accuracy with real RNA-seq data by comparing male and female LCL samples normalized with three different methods and comparing the results to a gold-standard set of genes with documented sex-specific expression. This assessment also shows that `tweeDEseq` provides a better precision-recall tradeoff than the compared NB-based methods (see Figure 10 and Additional file 2: Figure S11). We have also made a comparison with matching samples hybridised on microarray chips which allowed us to verify that `tweeDEseq` yields a higher degree of reproducibility of significance levels with respect to microrray data.

All these different comparative assessments have been performed against two of the most widely currently used methods for DE analysis of RNA-seq data, `edgeR` and `DESeq`, under four different parameter configurations each, since their default parametrisation is tailored towards very limited sample size. Making an informed decision on what is the most appropriate setup is not trivial for the non-expert user and, for this reason, it is important to underscore that `tweeDEseq` is competitive with all of the methodologies that follow from the different configurations of `edgeR` and `DESeq` without the need to set a single parameter.

The fact that the volcano plots from `tweeDEseq` and `limma`, derived from RNA-seq and microarray data, reveal that `limma` is able to find a larger number of DE genes from the gold-standard, suggests a long way still ahead of us to fully exploit the RNA-seq technology for DE. Not only regarding experimental aspects, but also statistical ones such as properly detecting and adjusting for unwanted sources of non-biological variability, for which there is currently no well-established available techniques, as in the case of microarray data.

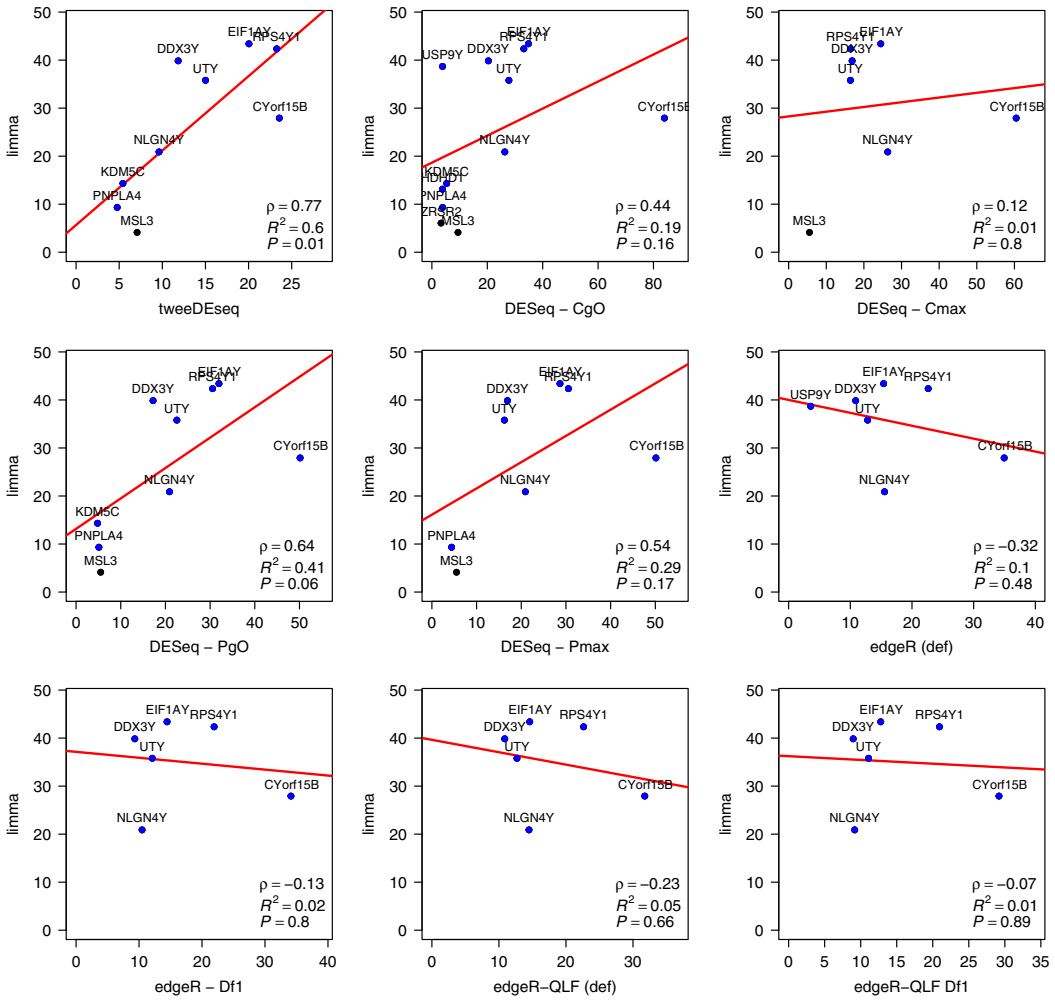

Figure 11 Reproducibility of differential expression (DE) between microarray and RNA-seq. Raw *P*-values of differential expression in $-\log_{10}$ scale for DE genes called at 10% FDR by both, limma (*y*-axis), from microarray data, and the other compared DE detection method applied on RNA-seq data (*x*-axis). A regression line is depicted in red. On the bottom-right corner of each panel, ρ indicates the Pearson correlation whereas R^2 and *P* indicate, respectively, the coefficient of determination and *P*-value of the test for zero regression coefficient, of the $-\log_{10}$ *p*-values of limma as function of those from the compared RNA-seq method. Only `tweeDEseq` provides a significant ($p < 0.05$) level of reproducibility between *P*-values of DE genes reported by both, limma on microarray data and the compared RNA-seq method, attaining also the highest ρ and R^2 values. Blue dots indicate genes with documented sex-specific expression.

Other applications of high-throughput sequencing technology that output counts of molecules, like in Copy Number Variation (CNV) analysis, could potentially benefit of models based on the PT-distribution. It is our perception that richer count data models of this kind will become increasingly necessary to draw accurate conclusions from data as technology brings us closer the actual biology of the cell.

Methods

Pre-processing of RNA-seq data

We have analyzed data from Pickrell *et al.* (2010) [12] that sequenced RNA from LCLs in 69 Nigerian (YRI) [12] individuals. Raw reads were downloaded from http://eqtl.uchicago.edu/RNA_Seq_data/unmapped_reads and pre-processed using the GRAPE pipeline [27]. This pipeline consists of first mapping the reads to the human genome

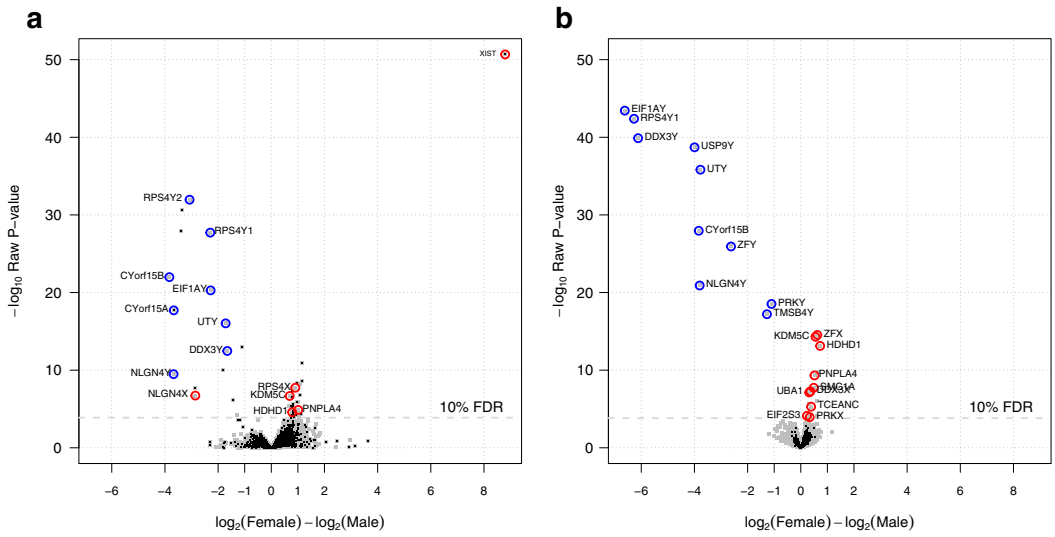

Figure 12 Comparison of DE analyses between microarray and RNA-seq. Volcano plots of DE analyses performed on matching LCL samples profiled with RNA-seq **(a)** and gene expression microarrays **(b)**. The *x*-axis corresponds to $\log_2$ fold-changes between female and male individuals while the *y*-axis corresponds to $-\log_{10}$ *P*-value of significance. RNA-seq data were analysed with `tweeDEseq` while microarray data were analysed with `limma`. Grey dots indicate genes common to both, the RNA-seq and the microarray gene expression matrices, while black dots indicate genes occurring exclusively in one of the two data sets. Blue and red circles indicate genes documented in the literature with sex-specific expression, concretely belonging to the male-specific region of chromosome Y and escaping X-chromosome inactivation in females, respectively.

version hg19 using the GEM mapper software [28]. Second, mapped reads were summarized into gene-level counts according to the GENCODE annotation version 3c [29] with Ensembl release 63 gene identifiers, by selecting those reads that mapped either completely within an exon or spanning a junction. This resulted in an initial table of counts of 38,415 Ensembl genes. This table of counts form part of the experimental data package `tweeDEseqCountData` available at http://www.bioconductor.org under the name `pickrell1`.

The table of counts was filtered to discard lowly expressed genes by keeping only those with an average of more than 0.1 counts per million (CPM) throughout the samples. The results shown in Additional file 2: Figure S11 were obtained by applying a more stringent minimum cutoff of 0.5 CPM. When we applied a normalization method that adjusted for gene length and G+C content (see below), genes without this information were also discarded. When the minimum CPM was 0.1, then 31,226 genes were kept when no normalization method or `edgeR`-TMM was applied and when `cqn` was applied then 27,438 were kept (see pg. 5 and 6 from Additional file 1). When the minimum CPM was 0.5 then these numbers decreased to 19,166 and 18,009 genes, respectively.

Three approaches to normalizing the table of counts from the LCL data have been considered. The first one is to work with the initial table of raw counts without any kind of normalization, the second one is to apply TMM [2] normalization as implemented in the `edgeR` [30] package, the third one is to use the methodology implemented in the `cqn` [4] Bioconductor package which adjusts for sample-specific effects of gene length and G+C content of every gene. When using the `DESeq` method for DE analysis in the LCL samples, the TMM normalization procedure was replaced by its own normalization procedure.

Raw counts were transformed into filtered and normalized counts for the purpose of producing MA-plots (Figure 2), assessing goodness of fit to the NB distribution (Figure 3), examining the relationship between mean expression level and the shape parameter of the PT distribution (Figure 4) and doing DE analysis with `tweeDEseq`. In the case of `DESeq` raw counts were transformed into normalized counts only when used with the `cqn` normalization method.

In the case of `edgeR`-TMM normalization, counts were transformed following the steps that the function `exactTest()` in `edgeR` takes: calculate normalization factors with the TMM method (`calcNormFactors()`), estimate effective library sizes and adjust counts to effective library sizes obtaining non-integer normalized pseudocounts (`equalizeLibSizes()`) which were

subtracted by 0.5 and then raised to the smallest integers not less than these pseudocounts (`ceiling()`). These steps are written together in the function `normalizeCounts()` from the `tweeDEseq` package.

In the case of `cqn`, normalization offsets are calculated by the function `cqn()` as $\log_2$ RPMs, which are added to original raw $\log_2$ RPMs. These are rolled back to absolute numbers and "unlogged" obtaining non-integer normalized pseudocounts which, analogously to the `edgeR`-TMM case, were subtracted by 0.5 and then raised to the smallest integers not less than these pseudocounts (`ceiling()`). The rationale behind subtracting 0.5 to the pseudocounts instead of directly truncating or raising to the next integer value, is to try to approach as much as possible the correct proportion of zero counts in the normalized data.

However, when performing DE analysis with `edgeR`, or with `DESeq` and its own normalization procedure, the specific recommendations made by the corresponding software authors were followed. More concretely, raw counts were not transformed in order to preserve their sampling properties and normalization adjustments entered the DE analysis through the corresponding normalization factors and offsets arguments within the functions that test for DE (see scripts for details in Additional file 1).

Pre-processing of microarray data

The microarray LCL data from [24] was processed from the raw CEL files available at http://www.ncbi.nlm.nih.gov/geo under accession GSE7792. Firstly, we only considered YRI samples. Secondly, data was processed using the Bioconductor oligo package. Quality assessment was performed by calculating NUSE and RLE diagnostics (Bolstad et al., 2005) and discarding those samples that either of the two reported diagnostics was considered below a minimum quality threshold. Third, using the RMA algorithm (Irizarry et al., 2003) implemented in the `rma()` function from the `oligo` package with argument `target="core"`, expression values were background corrected, normalized and summarized into Affymetrix transcript clusters. Fourth, most samples formed part of family trios and only samples belonging to father or mother were kept. Fifth, using the `getNetAffx()` function from the `oligo` package, Ensembl Transcript identifiers well obtained for each Affymetrix transcript cluster identifier. Sixth, using the bioconductor package `biomaRt`, Ensembl Transcript identifiers were translated into Ensembl Gene identifiers, resolving multiple assignments by keeping the Ensembl Gene identifier that had a match in the Ensembl Gene identifiers forming the table of counts of the [12] RNA-seq data, or choosing one arbitrarily, otherwise. Seven, duplicated assignments of the same Ensembl Gene identifier to multiple Affymetrix transcript cluster identifiers were resolved by keeping the transcript cluster with largest expression variability measured by its interquartile range (IQR).

At this point an expression data matrix of 16,323 Ensembl Genes by 74 samples was obtained and using the scanning date of each CEL file, samples were grouped into 5 batches, out of which one containing only three male samples was discarded leaving a total of 71 samples distributed into 4 balanced batches across gender. Batch effect was removed by using the QR-decomposition method implemented in the `removeBatchEffect()` function from the Bioconductor package `limma` [25] while keeping the sex-specific expression effect by setting the gender sample indicator variable within the design matrix argument. Finally, samples and genes were further filtered to match those from the RNA-seq table of counts.

Matching RNA-seq and microarray expression data matrices

To perform the analyses summarized in Figure 11 and Additional file 2: Figure S12 we further filtered the previously pre-processed RNA-seq and microarray gene expression matrices to match both Ensembl Gene identifiers and individual HapMap identifiers. This resulted in two gene expression data matrices of equal dimension with 15,194 genes and 36 samples. We only considered the RNA-seq data normalized with the `cqn` package.

To perform the analyses summarized in Figure 12 we built two other gene expression data matrices where, as before, samples were restricted to those 36 that matched between RNA-seq and microarray data but genes were not, leading to a RNA-seq and microarray gene expression data matrices of 27,438 and 16,323 Ensembl Genes by 36 samples, respectively. Genes were not matched since the purpose of these analyses was to gather insight into the differences and challenges in detecting DE genes using RNA-seq with respect to microarray gene expression data with many replicates.

Functional annotations

Functional annotations for Ensembl genes forming the tables of counts, were retrieved from `http://jun2011.archive.ensembl.org` with R and the `biomaRt` Bioconductor package. Gene length and G+C content annotations, used with the `cqn` normalization method, were obtained by downloading all human cDNAs from ftp://ftp.ensembl.org/pub/release-63/fasta/homo_sapiens/cdna/Homo_sapiens.GRCh37.63.cdna.all.fa.gz and calculating the length and G+C content of the longest cDNA for each Ensembl gene.

The gold-standard list of genes with sex-specific expression was built with genes reported in the literature that, in one hand, escape chromosome X inactivation [22] and, on the other hand, belong to the male-specific region of

chromosome Y [23]. In both cases, gene symbols were first translated into Ensembl gene identifiers and then further filtered to keep only those included in the set of Ensembl gene identifiers release 63. This resulted in a gold-standard list of 95 genes with sex-specific expression.

The list of housekeeping genes was retrieved from the literature [19] and mapped to Ensembl genes release 63, resulting in a final set of 669 housekeeping genes. The expression breadth reported in Figure 5 was obtained through the Barcode Gene Expression catalog [18] which uses information from 18,656 publicly available microarray samples from 131 tissue types, of the HG-U133 Plus 2.0 Affymetrix chip, to estimate the proportion of tissue types in which a given probeset is expressed in more than half the samples. After discarding unreliable probes (annotated with high-entropy in the catalog), we use these values as surrogates for expression breadth by mapping Affymetrix probeset identifiers to the genes in our table of counts through the `hgu133plus2.db` Bioconductor annotation package, leading to 16,292 genes with expression breadth values. When two or more probesets mapped to the same gene, the maximum value was taken for that gene.

All these functional data are included in the experimental data package `tweeDEseqCountData` available at http://www.bioconductor.org under the keywords annotEnsembl63, genderGenes and hkGenes.

Poisson-Tweedie distributions

Poisson-Tweedie (PT) distributions have been studied by several authors [31-34] and unify several over-dispersed count data distributions (see Figure one in [34]). This family of distributions can be defined by a probability generating function and mass probabilities have to be computed using a recursive algorithm [31,34]. El-Shaarawi *et al.* (2011) [34] compared different recursions and parameterizations of this family providing an algorithm to compute the PT probability distribution function. In the `R` package `tweeDEseq` we have developed a fast implementation, written in the C programming language, of this recursive algorithm. We briefly describe here the PT family of distributions as well as how we have used it to analyze RNA-seq count data in the context of a differential expression (DE) analysis.

Following El-Shaarawi *et al.* (2011) [34], let $Y \sim \text{PT}(a,b,c)$ be a PT random variable with vector of parameters $\theta = (a,b,c)^T$ defined in the domain

$$\Theta = (-\infty, 1] \times (0, +\infty) \times [0, 1)\,. \tag{1}$$

The PT random variable Y has a probability generating function (pgf) of the form:

$$G_Y(y|a,b,c) = \exp\left\{\frac{b}{a}\left((1-c)^a - (1-cy)^a\right)\right\}, \tag{2}$$

when $a \neq 0$, while when $a = 0$, then:

$$\lim_{a\to 0} G_Y(y|a,b,c) = \left[\frac{(1-c)}{(1-cy)}\right]^b. \tag{3}$$

Using this parameterization, the following recursive algorithm can be used to compute the PT probability distribution function [34]:

$$p_0 = \begin{cases} e^{b[(1-c)^a - 1]/a}, & a \neq 0, \\ (1-c)^b, & a = 0, \end{cases} \tag{4}$$

$$p_1 = bcp_0, \quad p_{k+1} = \frac{1}{k+1}\left(bcp_k + \sum_{j=1}^{k} j r_{k+1-j} p_j\right), \quad k = 1, 2, \ldots \tag{5}$$

where

$$r_1 = (1-a)c, \quad r_{j+1} = \left(\frac{j-1+a}{j+1}\right) c r_j, \quad j = 1, 2, \ldots \tag{6}$$

and p_i denotes the probability of observing i counts.

For the sake of interpretability, we reparameterize $\theta = (a,b,c)$ to $\theta = (\mu, \phi, a)$, where μ denotes the mean, $\phi = \sigma^2/\mu$ is the dispersion index (σ^2 is the variance), and a the shape parameter that is used to define some count data distributions that are particular cases of PT such as Poisson or negative binomial. The relationship between both parameterizations is the following:

$$c = \frac{\phi - 1}{\phi - a}, \quad b = \frac{\mu(1-a)^{(1-a)}}{(\phi-1)(d-a)^{-a}}\,. \tag{7}$$

The PT model includes not only Poisson ($a = 1$) and negative binomial (NB) ($a = 0$) but also other distributions that have been used to analyze count data such as Poisson-Inverse Gaussian (PIG) ($a = \frac{1}{2}$), Pólya-Aeppli (P-A) ($a = -1$) or Neyman type A ($a \to -\infty$). Therefore, the PT distribution family unifies several diverse count data distributions, including different overdispersed distributions such as NB or PIG. These distributions can model different scenarios as, for instance, a RNA-seq expression profile with a wide dynamic range leading to a heavy tail in the distribution. In such a case, PIG has a heavier tail than NB and this would make it more appropriate for such a gene. Note that an extremely heavy tail implies overdispersion, but the converse does not hold; hence the NB distribution is not adequate to model RNA-seq expression profiles of genes with a wide dynamic range due to their intrinsic biological variability [15].

Given a certain parameterization Kokonendji *et al.* (2004) [17] prove that the mean-variance relationship for the PT family can be expressed as:

$$\sigma^2 = \mu\left(1 + \mu^{p-1}\exp\left\{(2-p)\Phi_p\right\}\right) \tag{8}$$

where p is the shape parameter of that specific parameterization. It follows that, whereas the NB distribution is only able to capture a quadratic mean-variance relationship, the PT family is able to generalize this relationship to any order. As a result, it is more convenient to use the PT model when dealing with count data which presents variable overdispersion.

Parameter estimation for Poisson-Tweedie distributions

We need to estimate the parameter vector $\hat{\theta} = (\hat{\mu}, \hat{\phi}, \hat{a})$ to develop, on the one hand, a test of goodness-of-fit to an NB distribution and, on the other hand, a two-sample PT-test for differences in means. This latter test is used for detecting differentially expressed genes. Without loss of generality, let y_{gk} be the number of counts for gene g in sample k, derived from pre-processing RNA-seq data. We assume that y_{gk} follows the PT distribution:

$$y_{gk} \sim PT(\mu_g, \phi_g, a_g)\,. \tag{9}$$

In practice, we do not know the parameters $\theta_g = \mu_g, \phi_g, a_g$, but we can estimate them from data by maximum likelihood when the sample size is sufficiently large so that it guarantees the desirable large sample properties of unbiasedness and minimum variance of the maximum likelihood estimate (MLE). In the Additional file 2: Supplementary Information we provide a simulation study in order to estimate the minimum number of samples per group that approximately meets this requirement (see Additional file 2: Figure S7).

We obtained the MLE $\hat{\theta}$ using a quasi-Newton method with constraints. We have implemented such a procedure using the `optim` function in R. In order to guarantee good convergence, we consider as initial parameters the moment estimates of μ_g and ϕ_g, and $a_g = 0$. We choose this value for a_g because it corresponds to an NB model that is the natural cut-point of PT's parameter space.

Goodness-of-fit to a negative binomial distribution

In the framework of PT distributions we can formulate a test of the goodness of fit to an NB distribution by considering $H_0 : a = 0$ versus $H_a : a \neq 0$. Using a likelihood ratio test (LRT), the testing statistic is [34]

$$T = \frac{\max_{(\hat{\mu},\hat{\phi},\hat{a})} \ell(\hat{\mu}, \hat{\phi}, \hat{a} | y_0, \ldots, y_m)}{\max_{(\hat{\mu},\hat{\phi})} \ell(\hat{\mu}, \hat{\phi} | y_0, \ldots, y_m)}\,, \tag{10}$$

where numerator and denominator correspond to the likelihood functions for the PT and NB models, respectively. Since the PT model has just one parameter more than the NB model, the quantity $2 \log T \sim \chi^2_1$ under the null hypothesis, as n grows large, and it can be used to decide whether count data follow a NB distribution by means of a Q-Q plot (see Additional file 2: Figure S2) or by calculating the corresponding P-value.

Test to determine differentially expressed genes

For a given gene, let us assume that we observe $c_1, c_2, \ldots, c_n$ counts for n individuals and that we tabulate these counts into a contingency table with cells, $y_0, y_1, \ldots, y_m$ where $m = \max\{c_1, \ldots, c_n\}$. Therefore, y_c represents the number of observations with c counts. Then, the log-likelihood can be written as follows

$$\log \ell(\hat{\theta} | y_0, \ldots, y_m) = \sum_{i=0}^{m} y_i \, l_i(\hat{\theta})\,, \tag{11}$$

where $l_i(\hat{\theta}) = \log[\, p_i(\hat{\theta})]$ and $p_i(\hat{\theta})$ denotes the mass probability at i with $i = 0, 1, \ldots, m$ and is computed using the recurrence given in equation (6). El-Shaarawi *et al.* (2011) [34] indicate that when regularity conditions hold, that is, when θ is an interior point of the parameter space Θ, asymptotic normality of $\hat{\theta}$ can be assumed. Therefore, the negative inverse Hessian matrix of the log-likelihood at the MLE $\hat{\theta}$ corresponds to the estimated covariance matrix of $\hat{\theta}$. In particular, for the μ parameter we have that

$$\mathrm{Var}(\mu) = -\,\mathrm{E}\left[\frac{\partial^2}{\partial \mu^2} \log \ell(\hat{\theta} | y_0, \ldots, y_m)\right]^{-1}. \tag{12}$$

Consequently, if we are interested in comparing the mean counts for two sample groups, denoted by μ_A and μ_B, a two-sample PT-test for the mean with null hypothesis $H_0 : \frac{\mu_A}{\mu_B} = 1$, which we perform in logarithmic scale as $H_0 : \log(\mu_A) = \log(\mu_B)$, can be built by calculating the PT-statistic:

$$T = \frac{\hat{\mu}_A - \hat{\mu}_B}{\sqrt{\mathrm{Var}(\mu_A) + \mathrm{Var}(\mu_B)}}\,, \tag{13}$$

The PT-statitic, T, follows a standard normal distribution under the null hypothesis. Therefore, the $(1 - \alpha)$% percentile of a $N(0, 1)$ distribution is used to determine whether the observed differences between the two groups are statistically significant or not by providing a corresponding P-value that can be later on corrected for multiple testing using, for instance, Benjamini-Hochberg's FDR [35].

Simulation of RNA-seq data

The results shown in Figure 6 recreating the null hypothesis of no DE with real RNA-seq data were performed by dividing the LCL data into two separate data sets of male and female samples. From each data set we bootstrapped 100 times two groups of 20 samples uniformly at random, thus obtaining on the one hand, group pairs of female samples and, on the other hand, group pairs of male samples. On each bootstrapped data set we performed the two-sample test for DE detection of every method between the groups of female versus female samples and male versus male samples. We also considered two versions of the data, one with the raw un-normalized counts and the other with the counts normalized with the

cqn package [4]. In principle, there are no DE genes to be discovered from these comparisons, and therefore, under the null hypothesis of no DE, the *P*-value distribution for any given gene throughout the 100 bootstrapped data sets should be uniform.

The simulations shown in Figures 7, 8 and 9 contained synthetic RNA-seq data generated from a gamma-Poisson mixture model in a similar way to other published studies [8]. Under this model, we first draw dispersion parameters ϕ_g for every gene g at random from a gamma distribution Gamma($k = 2, \theta = 0.7$) and means according to three different fold-changes (1.5, 2 and 4) where half of the genes were up-regulated and the other half down-regulated. The λ_{gi} Poisson parameter for every gene g and sample i was drawn at random from a gamma distribution Gamma($k = a, \theta = 1/(\phi - 1)$) with $a = f\mu_{gk}/(\phi - 1)$ and $f \approx N(0, \sigma)$ corresponding to library factor which was either constant ($\sigma = 0$) or variable ($\sigma = 0.5$). Counts were simulated for each gene g from the resulting mixture gamma-Poisson distribution with parameters λ_{gi} for each sample i. Note that the resulting marginal distribution from the gamma-Poisson is a negative-binomial.

Software availability

- **Project name:** tweeDEseq
- **Project home page:** http://www.bioconductor.org/packages/release/bioc/html/tweeDEseq.html
- **Operating system(s):** Platform independent
- **Programming language:** R and C
- **Other requirements:** R 3.0.0
- **Licence:** GNU GPL
- Any restrictions tu use by non-academics: no restrictions

Additional files

Additional file 1: Scripts. ZIP file (.zip) containing all scripts, in the form of Sweave vignettes, to reproduce the results shown in this paper, including one copy of the resulting PDF file. Please read first through the README file contained in this tar ball in order to understand how to run the scripts.

Additional file 2: Supplementary materials. PDF file including supplementary figures and tables.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JRG and PP conceived the idea of modelling RNA-seq count data using PT family of distributions. ME programmed the recursive algorithm to compute PT probability distribution, performed simulation studies, and created the R package jointly with JRG and RC. PP and JRG proposed the statistical test for detecting DE genes. DG preprocessed the RNA-seq data. RC, ME and JRG analysed the data and wrote the paper. The project was supervised by JRG. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by grants from the 'Ministerio de Ciencia e Innovación - MICINN' (MTM2011-26515 to JRG and ME, MTM2010-09526-E to JRG, MTM2009-10893 to PP and TIN2011-22826 to RC), from European Reseach Council, Breathe project, ERC-AdG, GA num. 268479 to ME.

Author details

[1]Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. [2]Department of Mathematics, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. [3]Center for Genomic Regulation (CRG), Barcelona, Spain. [4]Department of Experimental and Health Sciences, Research Program on Biomedical Informatics (GRIB), Universitat Pompeu Fabra, Barcelona, Spain. [5]Hospital del Mar Research Institute (IMIM), Barcelona, Spain. [6]CIBER Epidemiology and Public Health (CIBERESP), Barcelona, Spain.

Received: 1 June 2013 Accepted: 14 August 2013
Published: 21 August 2013

References

1. Mortazavi1 A, Williams B, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5:**621–628.
2. Robinson M, Oshlack A: **A scaling normalization method for differential expression analysis of RNA-seq data.** *Genome Biol* 2010, **11:**R25.
3. Risso D, Schwartz K, Sherlock G, Dudoit S: **GC-content normalization for RNA-Seq data.** *BMC Bioinformatics* 2011, **12:**480.
4. Hansen KD, Irizarry RA, Wu Z: **Removing technical variability in RNA-seq data using conditional quantile normalization.** *Biostatistics* 2012, **13**(2):204–216.
5. Marioni J, Mason C, Mane S, Stephens M, Gilad Y: **RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18:**1509–1517.
6. Robinson MD, Smyth GK: **Small-sample estimation of negative binomial dispersion, with applications to SAGE data.** *Biostatistics* 2008, **9**(2):321–332.
7. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**(10):R106.
8. Lund SP, Nettleton D, McCarthy DJ, Smyth GK: **Detecting differential expression in RNA-sequence data using quasi-Likelihood with shrunken dispersion estimates.** *Stat Appl Genet Mol Biol* 2012, **11**(5). doi:10.1515/1544-6115.1826.
9. Hardcastle TJ, Kelly KA: **baySeq: empirical Bayesian methods for identifying differential expression in sequence count data.** *BMC Bioinformatics* 2010, **11:**422.
10. McCarthy DJ, Chen Y, Smyth GK: **Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation.** *Nucleic Acids Res* 2012, **40**(10):4288–4297.
11. Wu H, Wang C, Wu Z: **A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data.** *Biostatistics* 2012. doi:10.1093/biostatistics/kxs033.
12. Pickrell J, Marioni J, Pai A, Degner J, Engelhardt B, Nkadori E, Veyrieras J, Stephens M, Gilad Y, Pritchard J: **Understanding mechanisms underlying human gene expression variation with RNA sequencing.** *Nature* 2010, **464:**768–772.
13. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**(10):R80.
14. Van De Wiel MA, Leday GGR, Pardo L, Rue H, Van Der Vaart AW, Van Wieringen WN: **Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors.** *Biostatistics* 2012. doi:10.1093/biostatistics/kxs031.
15. Hansen K, Wu Z, Irizarry R, Leek J: **Sequencing technology does not eliminate biological variability.** *Nat Biotech* 2011, **29:**572–573.
16. Jorgensen B: *The Theory of Dispersion Models*. New York: Chapman and Hall; 1997.
17. Kokonendji C, Dossou-Gbété S, Demétrio C: **Some discrete exponencial dispersion models: Poisson-Tweedie and Hinde-Demétrio classes.** *SORT* 2004, **28**(2):201–214.
18. McCall M, Uppal K, Jaffee H, Zilliox R M J Irizarry: **The Gene Expression Barcode: leveraging public data repositories to begin cataloging**

the human and murine transcriptomes. *Nucleic Acids Res* 2011, **39:**D1011–D1015.
19. Eisenberg E, Levanon EY: **Human housekeeping genes are compact.** *Trends Genet* 2003, **19**(7):362–365.
20. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis.** *PLoS Genet* 2007, **3**(9):1724–1735.
21. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100**(16):9440–9445.
22. Carrel L, HF W: **X-inactivation profile reveals extensive variability in X-linked gene expression in females.** *Nature* 2005, **434:**400–404.
23. Skaletsky H, Kuroda-Kawaguchi T, Minx P, Cordum H, Hillier L, Brown L, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston R, Wilson R, Rozen S, Page D: **The male-specific region of the human Y chromosome is a mosic of discrete sequence classes.** *Nature* 2003, **423:**825–837.
24. Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME: **A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity.** *Proc Natl Acad Sci U S A* 2007, **104**(23):9758–9563.
25. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3:** doi:10.2202/1544-6115.1027.
26. Nguyen DK, Disteche CM: **Dosage compensation of the active X chromosome in mammals.** *Nat Genet* 2006, **38:**47–53.
27. Knowles DG, Röder M, Merkel A, Guigó R: **Grape RNA-Seq analysis pipeline environment.** *Bioinformatics* 2013, **29**(5):614–621.
28. Marco-Sola S, Sammeth M, Guigó R, Ribeca P: **The GEM mapper: fast, accurate and versatile alignment by filtration.** *Nat Methods* 2012, **9**(12):1185–1188.
29. Harrow J, Denoeud F, Frankish A, Reymond A, Chen CK, Chrast J, Lagarde J, Gilbert JGR, Storey R, Swarbreck D, Rossier C, Ucla C, Hubbard T, Antonarakis SE, Guigo R. *Genome Biol* 2006, **7**(Suppl 1):S4.1–S4.9.
30. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26:**139–140.
31. Hougaard P, Lee ML, Whitmore G: **Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes.** *Biometrics* 1997, **53:**1225–1238.
32. Gupta R, Ong S: **A new generalization of the negative binomial distribution.** *Compu Stat Data An* 2004, **45:**287–300.
33. Puig P, Valero J: **Count Data Distributions: Some Characterizations With Applications.** *J Am Stat Assoc* 2006, **101:**332–340.
34. El-Shaarawi A, Zhu R, Joe H: **Modelling species abundance using the Poisson-Tweedie family.** *Environmetrics* 2011, **22:**152–164.
35. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: A practical and powerful approach to multiple testing.** *J R Stat Soc B* 1995, **57:**289–300.

doi:10.1186/1471-2105-14-254
Cite this article as: Esnaola *et al.*: A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated RNA-seq experiments. *BMC Bioinformatics* 2013 **14**:254.

Summary of published papers

Introduction to the BREATHE project

Papers 1, 2 and 3 were developed as part of the BREATHE (BRain dEvelopment and Air polluTion ultrafine particles in scHool childrEn) project. Before summarising each of the papers contained in this thesis, we have deemed appropriate to first make a brief introduction to this project.

The BREATHE project was a longitudinal study funded by the European Research Council conducted from January 2012 to March 2013 in 39 schools in Barcelona (Catalonia, Spain). Its main research goal was to study the association between air pollution and cognitive development of schoolchildren. The study design considered 39 schools in Barcelona selected based on their traffic-related nitrogen dioxide (NO_2) values, pairing low and high NO_2 schools by socieconomic vulnerability index and type of school (i.e. public/private) to avoid residual social confounding. All school children (n = 5019) without special needs in grades 2 through 4 (7-10 years of age) were invited to participate and 2897 (59%) of them agreed.

Cognitive development was assesed through long-term change in working memory and attention. Children were evaluated every 3 months over four repeated visits using computerized tests. The computerized tests chosen were the n-back task on working memory, Anderson (2002), and the attentional network task (ANT), Rueda et al. (2004).

Exposure to traffic-related air pollution (TRAP) was measured in each pair of schools simultaneously twice during 1 week periods separated by 6 months. The pollutants measured were real-time concentrations of Black Carbon (BC), ultrafine particle number (UFP) concentration, particulate matter < 2.5 μm ($PM_{2.5}$) and NO_2. Exposure to TRAP at home was estimated using land use regression (LUR) models previously obtained as part of the ESCAPE project, Wang et al. (2013).

Socio-demographic factors were measured using a neighborhood socioeconomic vulnerability index based on level of education, unemployment and occupation in each census tract. Parents answered a questionnaire on family origin, gestational age and weight, breastfeeding, smoking during pregnancy and many other possibly confusing or mediating variables.

All things considered, it is important to note that the resulting data presented a complex structure. On the one hand, some variables had up to 4 repeated measurements (e.g. the cognitive development outcomes) while some others were measured just once (e.g. socio-economic status). On the other hand, we had both individually measured exposures (e.g. home TRAP)

and groupally measured ones (e.g. school TRAP).

For more information on the BREATHE project please visit the official web site of the project: https://breathe.isglobal.org/.

Summary of Paper 1

Paper 1 was developed as part of the BREATHE project (refer to *Introduction to the BREATHE project* for more information about it). In this paper we aimed to evaluate the psychometric properties and criterion validity of the n-back and ANT computerized tests used to assess the working memory and attention function, respectively. The main outcomes of the n-back test were d' scores (in broad terms, z scores of the difference between the correct and incorrect response rates) for 3 different difficulty levels and hit reaction time (HRT). The outcomes measured for ANT were incorrect responses, omissions, alerting, orienting and conflict.

Besides the listed test outcomes, other education and socio-demographic variables were considered: grade, sex, school performance, ADHD diagnosis, SDQ (Strengths and Difficulties Questionnaire) and maternal education.

Regarding the statistical analyses we first performed univariate descriptive statistics of the main outcomes and covariates. Then, we studied the pyschometric properties of both n-back and ANT tests using two different methods: Cronbach's alpha coefficient to study internal consistency and exploratory factor analysis (EFA) to study the factorial structure. To study the criterion validity of the n-back and ANT tests we studied the bivariate association between the outcomes and each preselected variables. Finally, the relationship between outcomes and the rest of covariates were also studied using Bayesian Networks (BN), Pearl (1985). From now on we will just focus on this last part, as we find it to be the most relevant from a statistical and methodological perspective.

A BN is a graphical model that describes compactly the dependency structure between a set of variables. It consists of a directed acyclic graph (DAG) that encodes the conditional dependencies between the variables and a set of local probability distributions associated to each of the variables. The variables are represented by the nodes of the DAG and the edges encode the conditional dependences. Our goal was to find the BN that best fitted our data in order to understand the complex interdependency structure between all considered variables. We calculated four DAGs, one for each n-back level of difficulty, including the d' scores for the four different stimuli, and one for ANT measures.

A BN can accommodate both discrete and continuous variables. However, in most cases the inference becomes intractable if we do not restrict to certain distributions. The most common approach, and our choice in

this paper, was to consider all variables to be discrete. Therefore, nodes can be modeled with multinomial distributions and Dirichlet (multivariate generalization of the beta distribution) priors. As our goal was to find the DAG that best describes our data, we first need some kind of score to order all possible BNs according to their goodness of fit to our data. We decided to use the K2 score, Cooper and Herskovits (1992), for its simplicity and appropriateness when dealing with discrete data.

When looking for the BN that best fits our data it is important to note that the number of possible DAGs grows super-exponentially ($2^{O(n^2)}$), Chickering et al. (2004)) as the number of variables (nodes) increases, thus making it unfeasible to calculate the score of all possible BNs. To solve this we used a Metropolis-Hastings random walk, Hastings (1970), over the whole DAG space. Moreover, to achieve greater validity, we decided to also use three alternative structure discovery methods: an order-based approach, Friedman and Koller (2000), the greedy hill-climbing method, Gamez et al. (2011), and exact Bayesian structure discovery, Koivisto (2006). All this was done using our own custom-made R package: *BayNet*, Esnaola (2013).

Figure 1 shows the four DAGs that resulted from applying Bayesian Model Averaging, Hoeting et al. (1999), to the posterior distributions obtained via the Metropolis-Hastings random walk over the DAG space. These results were validated when applying the other three alternative structure discovery methods. We observed that the three different DAGs for n-back showed similar structure: the four d' scores were strongly interrelated with one another. In addition, of all the preselected variables, we observed that these four scores were most strongly dependent on children's grade and school performance. Sex, maternal education, ADHD symptomatology, and behavioral problems were also associated with d' scores, though indirectly. The DAG for the ANT measures showed that the three attentional networks and incorrect responses were dependent on omissions. Similarly to the n-back DAGs, grade, and school performance showed the strongest dependencies with ANT measures. Sex, maternal education, ADHD symptomatology, and behavioral problems were also related to ANT measures, but less strongly. Additionally, child's sex and incorrect responses were robustly related.

The primary aim of our study was to determine whether the n-back and ANT are valid measures of child neuropsychological development for use in epidemiological studies. Our results demonstrated good criterion validity for the n-back d' scores by showing strong associations with a set of variables selected a priori based on previous knowledge. In addition, d'

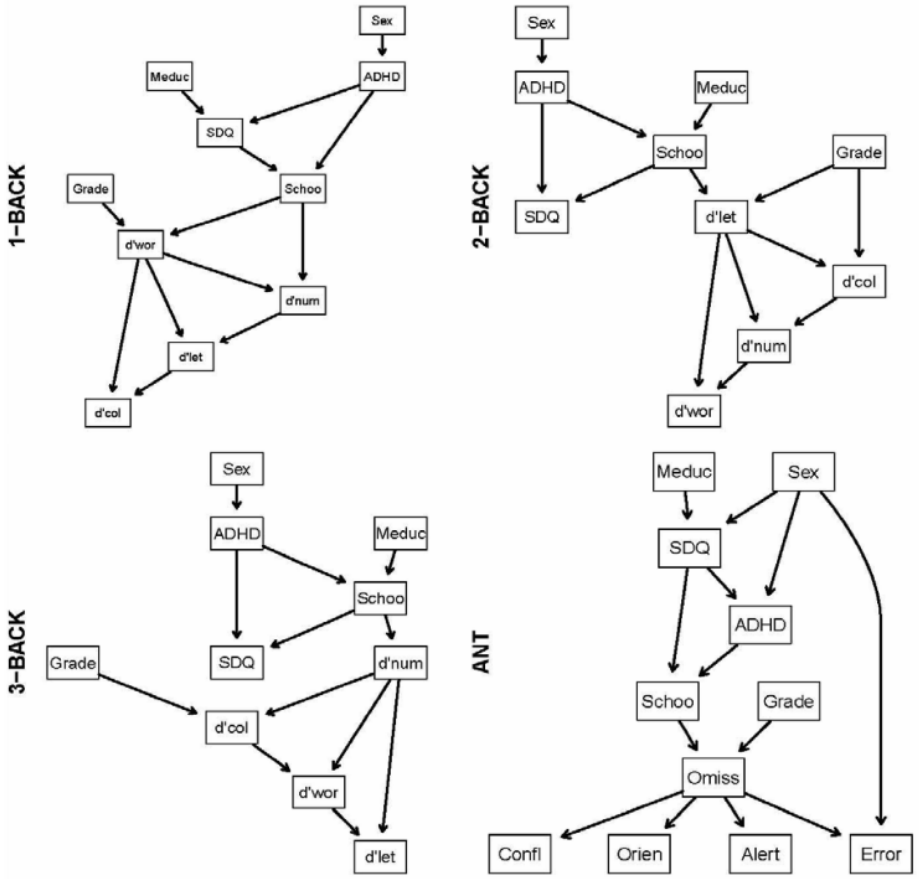

Figure 1: DAGs of the statistical dependencies between n-back and ANT outcomes and a set of preselected variables using Bayesian Networks.
Sex = Child's sex; ADHD = Attention Deficit and Hyperactivity Symptomatology; Meduc = Maternal education; SDQ = Strengths and Difficulties Questionnaire; Schoo = School performance; Grade = Child?s grade; d'wor = d'score for words; d'num = d'score for numbers; d'let = d'score for letters; d'col = d'score for colors; Omiss = omissions; Error = Incorrect responses; Confl = Conflict Score; Orien = Orienting Score; Alert = Alerting Score.

scores also showed local and global statistical dependencies with all of these preselected variables. With regard to the ANT, we observed that the most sensitive outcomes were incorrect responses, omissions, and the conflict score. Direct measures of error (incorrect responses and omissions) showed consistent associations with all preselected variables such as age, sex, school performance, maternal education, as well as with clinical measures of ADHD clinical criteria and behavioral problems.

Our results show that both the n-back and ANT are valid and relatively easy-to-apply tests for measuring child neuropsychological development in epidemiological studies. Most of the measures obtained in the n-back (d') and the ANT (incorrect responses, omissions, and conflict score) associated with working memory and attention capacities showed good psychometric properties, good criterion validity, and acceptable statistical dependencies. Thus, we recommend their use in large epidemiological studies.

Summary of Paper 2

Paper 2 was developed as part of the BREATHE project (refer to *Introduction to the BREATHE project* for more information about it). In it we studied the association between Traffic Related Air Pollution (TRAP) and child neurodevelopment.

Air pollution is a suspected developmental neurotoxicant, Grandjean and Landrigan (2014). In children, exposure to traffic-related air pollutants during pregnancy or infancy, when the brain neocortex rapidly develops, has been related to cognitive delays, Guxens et al. (2012). Children spend a large proportion of their day at school, including the period when daily traffic pollution peaks. Considering all this, our goal was to assess the relationship between long-term exposure to traffic-related air pollutants at school and cognitive development measurements in primary school children.

As outcomes of our models, for the n-back test we selected two-back and three-back loads for number and word stimuli as they showed a clear age-dependent slope in the four measurements and had little learning effect. The two-back test predicts general mental abilities, while the three-back test also predicts superior functions such as fluid intelligence.

Regarding the exposures, outdoor and indoor long-term school air pollution levels were considered. To achieve a better spatial long-term average, elemental carbon (EC) and nitrogen dioxide (NO_2) were adjusted for temporal variability. We decided not to seasonalize ultrafine particles (UFP) because seasonalized UFP had a poorer correlation between the two measurement campaigns than non-seasonalized UFP.

A total of 2715 children with complete data (i.e. repeated outcome at least twice and individual data on maternal education and age) were included in the study. In total, we had 10112 outcome observations (3.72 test outcomes per individual on average). Because of the multilevel nature of the data (i.e., visits within children within schools), we used linear mixed-effects models with the cognitive parameters (test performance) from the four repeated visits as outcomes and random effects for child and school. Age (centered at visit 1) was included in the model in order to capture the growth trajectory of cognitive test performance. An interaction between age at each visit and school air pollution was included to capture changes in growth trajectory associated with school air pollution exposure. The main effect of air pollution (AP), which was also included in the model, captures the baseline differences in cognitive function that are associated

with air pollution. This model was further adjusted for potential confounders selected with directed acyclic graphs. The final adjusted model included additional coefficients for sex, maternal education , residential neighborhood socioeconomic status, and air pollution exposure at home:

$$Y_{sit} = \beta_0 + \beta_1(\text{Age}_t - \text{Age}_1) + \beta_2\text{AP} + \beta_3(\text{Age}_t - \text{Age}_1)\text{AP} + \beta_4\text{Sex}$$
$$+\beta_5\text{Mat\_educ\_primary} + \beta_6\text{Mat\_educ\_secondary} + \beta_7\text{Mat\_educ\_university}$$
$$+\beta_8\text{Neighborhood\_socioeconomic\_status} + \beta_9\text{Air\_pollution\_exposure\_at\_home}$$
$$+u_s + v_{i(s)} + \varepsilon_{sit}$$

where Y_{sit} represents the cognitive test result for subject i in school s at visit t, t = $\{1, 2, 3, 4\}$; u_s are random effects at school level, assumed to be normally distributed with expectation 0 and variance σ_u^2; $v_{i(s)}$ are random effects associated with subject i in school s, assumed to be normally distributed with expectation 0 and variance σ_v^2; and ε_{sit} are the model residuals, assumed to be normally distributed with expectation 0 and variance σ_ε^2. All models were fitted using the *lme4* R package, Bates et al. (2015).

Table 1 shows the adjusted air pollution coefficients at baseline and per 12-month change for all the studied cognitive outcomes. Children attending schools with higher levels of EC, NO_2 , and UFP both in the courtyard and in the classroom had worse cognitive parameters at baseline than children attending schools with lower air pollution. All the coefficients were negative for working memory and positive for inattentiveness, indicating impairment, though the differences were not statistically significant. The growth in cognitive parameters during the 1-year follow-up was also reduced in the schools exposed to higher air pollution levels, which in consequence amplified the differences between schools at the end of follow-up. The detrimental association of air pollution with change in the cognitive parameters was observed for all the outcomes and pollutants, being statistically significant for almost all of them.

Figure 2 shows the change in working memory in 1-year as a function of both outdoor and indoor pollutant levels. The points in the figure represent the crude estimates of change in cognitive parameters for each school along with the school air pollution levels, while the line represents the regression line obtained from the final adjusted model. This figure illustrates the negative relationship between change in cognitive function and air pollution levels, and depicts a good fit between the crude values and the adjusted slope.

Cognitive Outcome	High/Low Traffic	Outdoor (Courtyard)			Indoor (Classroom)		
		EC	NO_2	UFP	EC	NO_2	UFP
Units per interquartile range	—	0.7 μ g/m^3	23.3 μg/m^3	6,110 counts	0.92 μg/m^3	18.1 μg/m^3	8,872 counts
Working memory (two-back numbers, *d'*)							
Baseline	−5.3 (−16, 5.1)	−5.8 (−12, 0.56)	−7.5 (−16, 0.99)	−6.4 (−14, 1.5)	−3.0 (−11, 4.8)	−6.1 (−14, 1.9)	−1.3 (−13, 9.9)
12-mo change	−9.9 (−16, −3.5)*	−4.1 (−8.0, −0.2)*	−6.6 (−12, −1.2)*	−4.9 (−10, 0.22)	−6.2 (−11, −2.0)*	−5.6 (−11, −0.44)*	−7.9 (−15, −1.3)*
Superior working memory (three-back numbers, *d'*)							
Baseline	−1.4 (−10, 7.1)	0.25 (−5.2, 5.7)	1.5 (−5.8, 8.8)	−0.95 (−7.4, 5.6)	1.4 (−5.0, 7.9)	1.3 (−5.4, 8.0)	−0.078 (−9.1, 8.9)
12-mo change	−5.8 (−11, −0.74)*	−4.4 (−7.6, −1.3)*	−6.7 (−11, −2.3)*	−5 (−9.1, −0.96)*	−5.8 (−9.2, −2.4)*	−5.1 (−9.2, −0.91)*	−6.0 (−11, −0.75)*
Inattentiveness (HRT-SE, milliseconds)							
Baseline	5.2 (−6.2, 17)	1 (−6.3, 8.4)	4.8 (−5.0, 14)	4.5 (−4.0, 13)	6.8 (−1.7, 15)	7.0 (−1.8, 16)	6.2 (−5.8, 18)
12-mo change	5.2 (0.68, 9.7)*	3.8 (1.0, 6.6)*	3.8 (−0.10, 7.6)	3.9 (0.31, 7.6)*	3.8 (0.79, 6.8)*	2.6 (−1.0, 6.3)	4.6 (−0.13, 9.2)

Table 1: Difference (95% CI) in the 12-month change adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects. (*p ¡ 0.05).

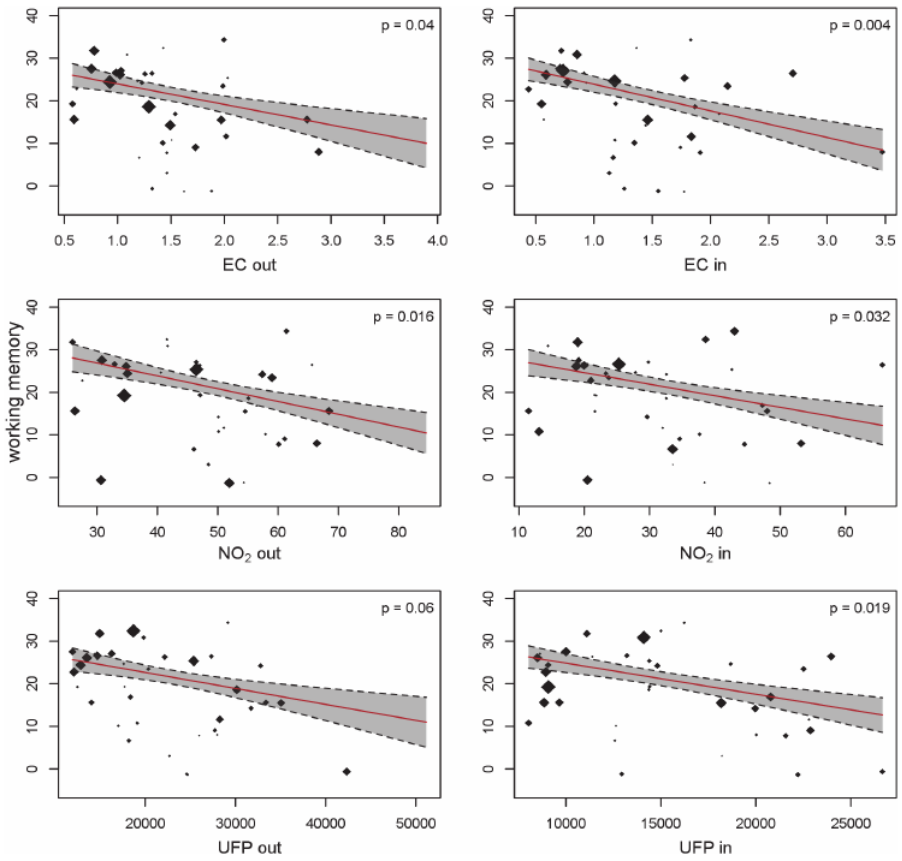

Figure 2: Working memory development and long-term exposure to traffic-related air pollutants. Each dot depicts a school, with size proportional to the number of children. The cognitive development per school was estimated in a model with school and individual as random effects. The slope of the red line depicts the change in cognitive development as a function of the air pollutants, adjusted for age, sex, maternal education, residential neighborhood socioeconomic status, and air pollution exposure at home; school and individual as nested random effects in 2715 children and 10112 tests from 39 schools. Gray shading indicates 95% CIs. out, outdoors (courtyard); in, indoors (classroom).

We performed sensitivity analyses to rule out posible confounding due to not included variables. We further adjusted our models for other individual socio-economic factors, ADHD or behavioral symptoms, residential greenness, and school noise. Results remained unchanged, indicating that our results are robust. In stratified analyses, stratifying for high/low TRAP schools, neighborhood socio-economic status and obesity did not alter our results. In contrast, detrimental associations were stronger in general in boys than in girls, in children from more highly educated mothers, in children from private schools, and in children with ADHD symptoms, though differences were not significant. All this considered, we concluded that our findings were robust.

Our study demonstrated that cognitive development is reduced in children exposed to higher levels of traffic-related air pollutants at school. This association was consistent across all considered neurodevelopment outcomes. The findings provide strong support for air pollution being a developmental neurotoxicant and point towards the primary school age as a particularly vulnerable time window for executive function development. Overall, we have shown that children attending schools with higher levels of exposure to traffic-related air pollutants had a smaller growth in cognitive development over time, suggesting that traffic-related air pollution in schools negatively affects cognitive development. This may have consequences for learning, school achievement, and behavior. With regard to air pollution regulation, the present study shows that the developing brain may be vulnerable to certain traffic-related air pollutants.

Summary of Paper 3

Paper 3 was developed as part of the BREATHE project (refer to *Introduction to the BREATHE project* for more information about it). In it we studied the role of PM2.5 (particulate matter of diameter $< 2.5\mu m$) sources in school air on cognitive development.

In paper 2 we reported that cognitive development over 1 year showed a slower increase among children attending schools with high traffic-related air pollution levels compared with children in less polluted schools. In that study, the air pollution markers used were nitrogen dioxide (NO_2), elemental carbon (EC), and ultrafine particle (UFP) number. Interestingly, PM2.5 mass concentrations at the studied schools were not correlated with traffic air pollution, and most of the contribution to PM2.5 levels was due to mineral and organic sources. PM2.5 levels are the universal indicator of air quality because of their overwhelming adverse association with many health indicators (WHO 2013). In this paper, we aim to explore the role of all the different sources of PM2.5 in school air on cognitive development.

Source	Identifying species (tracers)
Mineral	Al, Mg, Li, Fe, Ca, Ti, Rb
Traffic	EC, Cu, Sb, Sn, Fe
Organic/textile/chalk	OC, Ca, Sr
Secondary sulfate and organics	SO4-, NH4+
Secondary nitrate	NO3-
Road dust	Ca, Fe, Cu, Sb
Metallurgy	Zn, Pb, Cd, Mn, Cu
Sea spray	Na, Cl-
Heavy oil combustion	V, Ni

Table 2: Main elements identifying the estimated sources.

We conducted a source apportionment analysis to estimate the levels of the different components present in PM2.5. This was performed using a constrained positive matrix factorization (PMF) model based on 33 chemical species. PMF is a weighted least squares technique that allows accounting for the uncertainty associated with the analytical procedure. This technique

returned a solution that identified nine main factors/sources responsible for the variability of PM2.5 mass concentrations with an R^2 of 0.95. The nine identified sources together with their elemental composition are shown in table 2.

Similarly to the analyses in paper 2, we used linear mixed-effects models with the four repeated cognitive parameters as outcomes and random effects for child and school to accommodate the multilevel nature of the data (i.e. visits within children within schools). Age at each visit (centered at visit 1) was included in the model to capture the growth trajectory of the cognitive test. An interaction between age and school concentrations of individual PM sources was included to capture changes in growth trajectory associated with school air pollution exposure. The latter was the effect of interest in this study. Potential confounders were identified using directed acyclic graphs (DAG) and they included sex, maternal educa- tion (primary or less/secondary/university), residential neighborhood socioeconomic status, and air pollution exposure at home. Indicators of school pair were included in the model to restrict comparisons within pairs of schools measured during the same days, thus removing potential differences in air pollution levels between schools that were attributable to meteorology or seasonality. The model equation was the following,

$$\begin{aligned} Y_{psit} = \beta_{0p} + \beta_1(\text{Age}_{psit} - \text{Age}_{psi1}) + \beta_2(\text{PM2.5\_source})_{ps} \\ +\beta_3(\text{Age}_{psit} - \text{Age}_{psi1}) \times (\text{PM\_source})_{ps} \\ +\mathbf{Z}\boldsymbol{\eta} + \text{u}_{ps} + \text{v}_{psi} + \varepsilon_{psit} \end{aligned}$$

where Y_{psit} is the cognitive test result for subject i in school s (belonging to pair p) at visit t, β_{0p} are pair-specific intercepts, $\mathbf{Z}$ is a matrix including all confounders, $\boldsymbol{\eta}$ is a vector of parameters associated to confounders, u_{ps} are random effects at school level, assumed normally distributed with expectation 0 and variance σ_u^2, v_{psi} are random effects associated with subject i in school s, assumed normally distributed with expectation 0 and variance σ_v^2, and ε_{psit} are the model residuals assumed normally distributed with expectation 0 and variance σ_e^2.

Figure 3 displays the change in cognitive outcomes over the follow-up period for an interquartile range increase in source-specific PM 2.5 concentrations. Changes from the first to the third quartile in the indoor traffic source were associated with a significant reduction in working memory of -5.6 [95% confidence interval (CI): -10.7, -0.5], equivalent to 22% of

the annual change experienced by the participants; a reduction of superior working memory of -5.1 (95% CI: -9.2, -1.1), equivalent to 30% of the annual change; and an increase of 3.6 (95% CI: 0.0, 7.1) in inattentiveness scale, equivalent to 11% of the annual change.

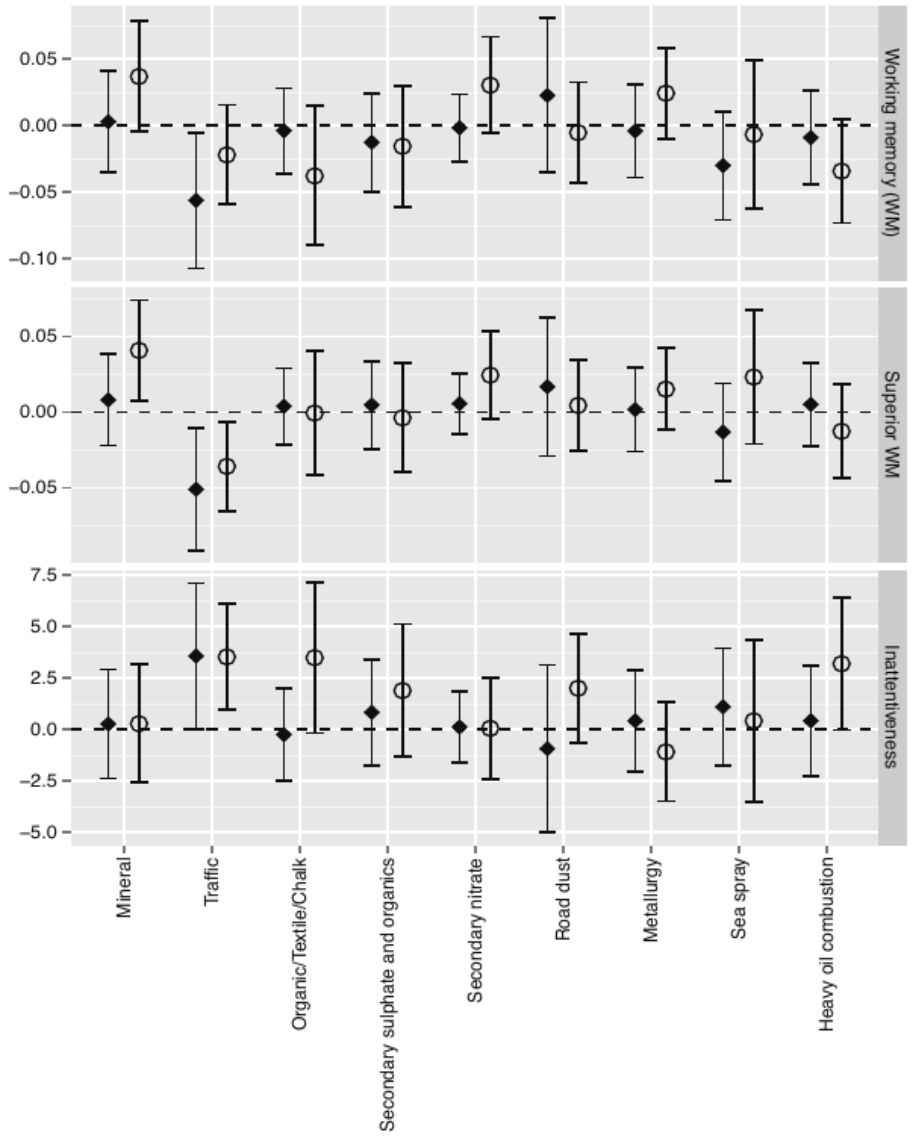

Figure 3: Change (95% CI) in cognitive growth per interquartile range increase in school source-specific PM 2.5 mass concentrations. Black diamonds: indoor concentrations; open circles: outdoor concentrations.

Summarising the results, we found that children attending schools with high levels of traffic-related PM 2.5 showed a slower cognitive development. None of the other PM 2.5 sources (mineral, organic/textile/ chalk, sulfate, nitrate, road dust, metallurgy, and sea spray) showed a deleterious association with cognitive development, although associations for heavy oil combustion were also suggested. Future studies should examine whether the effects observed at primary school age are long-lasting and have consequences over the life course.

Summary of Paper 4

In paper 4 we presented our method for the analysis of expression profiles arising from extensively replicated RNA-seq experiments.

RNA-seq is a technique to quantify the amount of RNA in a biological sample by direct ultra-high-throughput sequencing of cDNA (also known as Next Generation Sequencing). The resulting sequence reads are individually mapped to the source genome and counted to obtain the number and density of reads corresponding to RNA from each known exon, splice event or new candidate gene. One of the main advantages of RNA-seq over previous hybridization-based microarray technology is a much larger dynamic range of detection. However, the extent to which this feature is fully exploited depends entirely on the way the resulting data is analyzed when addressing a particular biological question. For instance, in the identification of genes that significantly change their expression levels between groups of samples, also known as differential expression (DE).

Detection of DE genes using RNA-seq data was firstly based on using models assuming a Poisson distribution, which assumes that the mean and variance of the distribution are identical. Given that the observed variation in read counts is much larger than the mean (overdispersion), researchers have proposed the use of negative binomial (NB) distributions which are defined by two parameters: the mean and the dispersion. However, the larger power of RNA-seq to capture biological variability can potentially introduce into count data not only overdispersion, but also oddities such as zero-inflation (i.e., in lowly expressed genes, the proportion of zero counts may be greater than expected under an NB distribution) and heavy tail behavior (i.e., a large dynamic range within the same expression profile), especially when many biological replicates are available. Under these circumstances even a two-parameter NB distribution may not provide an adequate fit to the data (see Figure 4). In turn, this may lead to incorrect statistical inferences resulting in lists of DE genes with a potentially increased number of false positive calls and poor reproducibility. To overcome this problem, methods based on the NB distribution use sophisticated moderation techniques that borrow information across genes and exploit the mean-variance relationship in count data to improve the estimation of the NB dispersion parameter.

In this paper we propose to approach this problem by using other count data distributions that fit expression profiles better than the NB without the need to alter configuration parameters. We illustrate how a more flexible

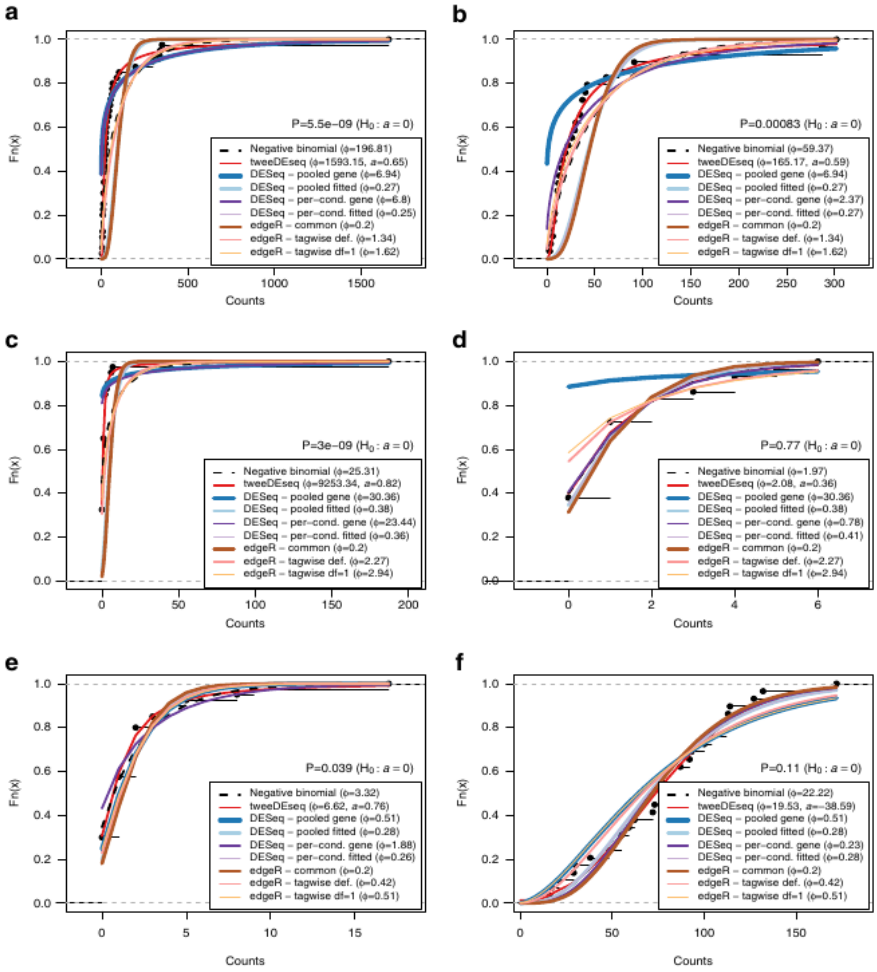

Figure 4: Fit of different count data distributions to diverse real RNA-seq gene expression profiles. All plots show the empirical cumulative distribution function (CDF) of counts (black dots) and the CDF estimated by a pure negative binomial model (black dashed line), a Poisson-Tweedie model (red line) obtained with tweeDEseq and several moderated negative binomial models obtained with different parameter configurations of DESeq and edgeR. Estimated dispersions, and shape in the case of tweeDEseq, are indicated in the legend. Above the legend, the P-value of the test of goodness-of-fit to a negative binomial distribution is shown.

family of count-data probability distributions, called the Poisson-Tweedie, provides a better fit to these expression profiles. We then introduce a new test for differential expression analysis in RNA-seq data based on the Poisson-Tweedie family of distributions. By surveying the tiny fraction of sex-specific gene expression changes in LCL samples, we approach the problem of assessing accuracy in DE analysis with real RNA-seq data and show that, in the context of extensively replicated RNA-seq experiments, our method yields better performance than competing NB-based methods (edgeR and DESeq).

Poisson-Tweedie (PT) distributions unify several over-dispersed count data distributions. This family of distributions can be defined by a probability generating function and mass probabilities have to be computed using a recursive algorithm. In the R package tweeDEseq we have developed a fast implementation, written in the C programming language, of this recursive algorithm. The PT model includes not only Poisson and negative binomial (NB) but also other distributions that have been used to analyze count data such as Poisson-Inverse Gaussian (PIG), Plya-Aeppli or Neyman type A. These distributions can model different scenarios as, for instance, a RNA-seq expression profile with a wide dynamic range leading to a heavy tail in the distribution. For the purpose of a DE analysis between two groups of samples, we have developed a two-sample PT-test for differences in means.

Figure 4 illustrates the flexibility of the PT distribution to accurately fit different gene expression profiles. Each plot shows the empirical cumulative distribution of observed counts as well as the parametric cumulative distributions obtained through the estimation of parameters of the methods compared in this paper under different configurations. As it can be observed, PT is able to correctly capture the distribution in each of the 6 cases, whereas methods based on the NB (edgeR and DESeq) fail to do so in some of the cases.

Regarding the results of the DE analyses, Figure 5 shows the precission and recall comparison on the LCL RNA-seq data. As it can be seen, tweeDEseq provides higher F-measure values than the other two NB based methods, thus indicating a better precission-recall tradeoff.

Summarising, the increased amount of biological variability revealed by extensive replication in RNA-seq experiments brings new challenges to the task of identifying genes whose change in expression is both, biologically and statistically significant. In this paper we propose using the PT family of distributions, showing that it captures a much richer diversity of expression dynamics in RNA-seq count data than the statistical models based in the NB

distributions alone.

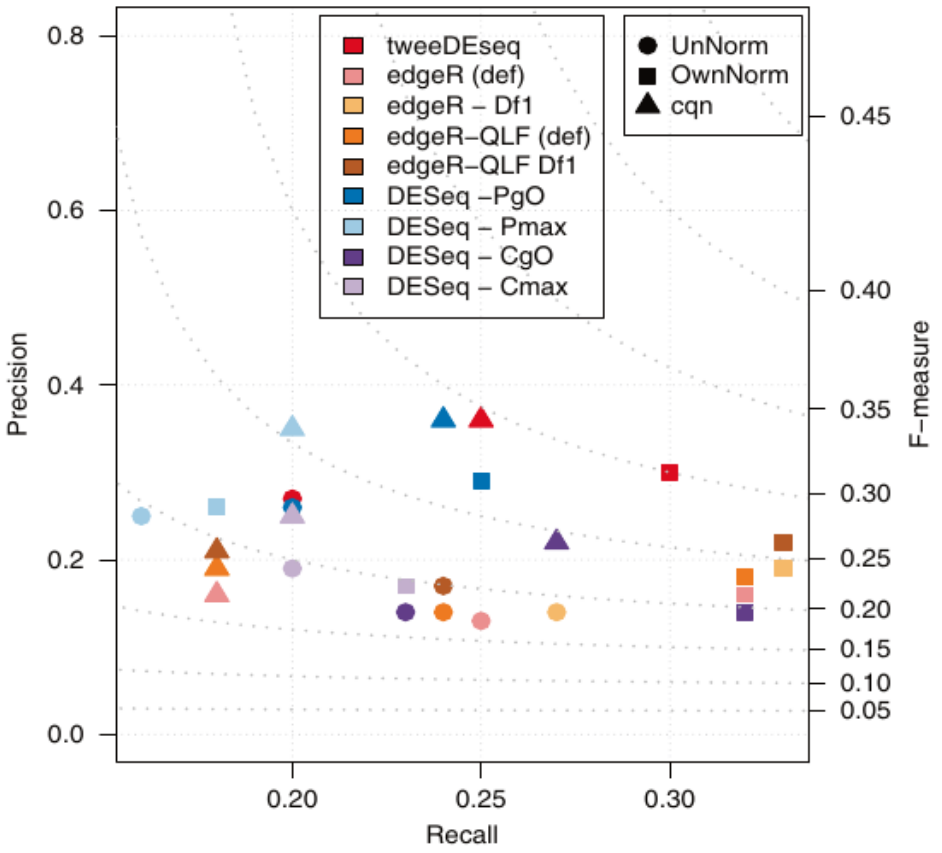

Figure 5: Precision and recall comparison on the LCL RNA-seq data. Precision (y-axis) and recall (x-axis) values for genes called DE at 1% FDR by different DE detection methods and configuration parameters. The right y-axis indicates values of the F-measure shown by dot lines.

Conclusions

Recent advances in epidemiology are driving the development of new statistical methodology. In this thesis we have presented a compendium of published papers dealing with the application and development of some of these methods. We will first discuss each paper's specific conclusions and then move into more general conclusions.

Conclusions to Paper 1

In paper 1 we applied Bayesian Networks to learn about the complex structure of interdependencies between a set of variables comprising both sociodemographic and neurodevelopmental proxies. The obtained results allowed to demonstrate that both the n-back and the ANT are valid and relatively easy-to-apply tests for measuring child neuropsychological development in epidemiological studies. Thus, the use of Bayesian Networks proved to be of great usefulness for the posterior statistical analyses as it allowed us to detect mediating and confusing variables. Regarding the validity of our results, it must be noted that, due to the vast number of possible DAGs, there is no proper way to assess the optimality of the obtained results. However, having obtained almost identical results when using three alternative and fundamentally different search methods proves that our results are reliable and robust. We recommend this approach when studying the validity of the results in future Bayesian Network structure search analyses.

As for the limitations of the analysis, it should be emphasized that all considered variables had to be discrete. As a result, continuous variables

had to be discretized before conducting the analysis, thus losing potentially significant information. This could be circumvented by expanding the probabilistic structure of the considered Bayesian Networks. For instance, instead of using the multinomial distribution for modeling each node's conditional dependencies, one could use a mixture of discrete and continuous distributions. The problem with this approach is that it adds a significant amount of necessary computations to each iteration of the structure search algorithms, thus rendering them impractical for a moderate number of variables (nodes).

The use of BNs poses many challenges due to the associated methodological and computational complications. However, the rapid development of computing technology together with the refinement of current estimation methodologies are allowing to apply BNs in many other settings such as OMICS pathway analysis, image processing or decision making. The versatile nature of BNs together with their ability to capture both local and global variable inter-relationships offer unprecedented opportunities in the modeling of complex data, thus giving them the potential to revolutionize many research fields.

Conclusions to Papers 2 and 3

In paper 2 we studied whether exposure of children in primary school to traffic-related-air pollution is associated with impaired cognitive development. Our results showed that cognitive development is reduced in children exposed to higher levels of traffic-related air pollutants at school. These findings were consistent across the different analysed pollutants. Nevertheless, it should be noted that residual confounding for social class could not be fully discarded, even when adjusting for neighborhood socio-economical index, maternal education and other similar socio-economical proxies. However, found associations remained in stratified and sensitivity analyses, thus rendering highly unlikely the possibility of residual confounding.

In paper 3 we explored the role of PM sources in school air on cognitive development. We found that, among all the different PM sources, traffic was the only one associated with a reduction in cognitive development. We also found an association between outdoor levels of heavy oil combustion and two of the outcomes but chance could not be excluded due to the lack of association when analysing Nickel and Vanadium, the main elements defining this source. Unexpectedly, we also found that exposure to mineral particles was beneficial for superior working memory. This could be a

chance finding, but it could also be explained by the fact that schools with higher mineral concentrations have more greenness, which has already been demonstrated to be beneficial for cognitive development in school children. All these results are in line with our findings in paper 2.

The methodology in both papers 2 and 3 heavily rely on Linear Mixed-effects Models. Not only are LMMs able to satisfactorily capture the multilevel nature of the data, but they also have the capacity to incorporate complex correlation structures. These two things are key aspects in current and future epidemiologic studies, which are increasingly growing in complexity.

Conclusions to Paper 4

In paper 4 we presented `tweeDEseq`, a method for the analysis of RNA-seq data based on the Poisson-Tweedie family of distributions. Using both real and simulated data, we showed that, in the context of extensively replicated RNA-seq experiments, our method outperforms mainstream competing methods based on the Poisson or Negative-Binomial distributions. Having three parameters, the Poisson-Tweedie family of distributions is able to accommodate many of the diverse oddities arising from these experiments such as zero-inflation or heavy-tail behaviour. Conversely, methods based on the Poisson or Negative-Binomial distributions struggle when adjusting for such oddities due to their lack of flexibily. We also assessed our method's reproducibility by comparing its results against those obtained from microarrays. `tweeDEseq` is openly available as an `R` package in the `Bioconductor` repositories. Regarding our method's limitations, it should be noted that there is no closed form for the density function of the Poisson-Tweedie family of distributions. As a result, the parameter estimation process is not trivial and might present problems when one or more parameters are close to their boundary.

Since our paper was published, our method has been compared against other state of the art RNA-seq methods. Tang et al. (2015) evaluated several methods for differential expression analysis on multi-group RNA-seq count data. In it, they mention that "the Poisson-Tweedie model well captures the biological variation (especially for zero-inflation and heavy tail behavior) when many biological replicates are available", thus concluding that "Poisson-Tweedie may be the first choice four count data with many biological replicates".

The use of the PT family of distributions can have a great impact in

the short and mid term, as the amount of biological variability increases due to the use of extensively replicated RNA-seq experiments. Also, other applications of high-throughput sequencing technology that output counts of molecules, like in Copy Number Variation (CNV) analysis, could potentially benefit of models based on the PT family of distributions. We believe that richer count data models of this kind will become increasingly necessary to draw accurate conclusions from data as technology brings us closer to the actual biology of the cell. Furthermore, it should be noted that the possible applications of the PT family of distributions are not limited to the medical and health sciences. Other research fields such as econometrics or geology can greatly benefit from its flexibility modeling count data.

General conclusions

Overall, the present work has dealt with some of the main issues to consider in the design and analysis of modern epidemiologic studies. Firstly, we have shown how helpful Bayesian Networks can be in the design and first steps of a statistical analysis, as they allow to have a wide picture of the interdependencies between a set of variables. Secondly, we have demonstrated the power of Mixed-effects models when dealing with complex hierarchichal data. Finally, we have presented our method for the analysis of RNA-seq data, which is encompassed within the analysis of OMICs in the context of epidemiological studies. However, one should bear in mind that, far from being solved, many of the challenges arising from new epidemiologic studies are still to be fathomed.

Future work

In this thesis we have introduced statistical methodology and application for the analysis of complex epidemiology data. More specifically, we have presented applications of Bayesian Networks, Mixed-effects models and the Poisson-Tweedie family of distributions to particular scenarios arising in current epidemiology studies. Although this manuscript shows the usefulness and appropriateness of such methods, many challenges still need to be overcome. This section discusses possible future work to address some of these issues.

Bayesian Networks

Bayesian Networks are a powerful tool to visualize and model complex relationships. Although this method has already proven its value in a wide range of applications it still poses many challenges, mainly due to its associated computational complexity. Many solutions have been proposed to overcome these difficulties (for instance, the previously explained Order MCMC) but it still unfeasible for current methodologies to accommodate a moderate number of variables with the possibility of having different underlying propability distributions.

Recent and expected future advances in computational technology offer a promising way to tackle these issues. Nonetheless, we believe that instead of blindly relying on future technological progress, the focus should be put on the improvement of present-day methodologies. As most of the computational problems occur during the calculation of the exact analytical

joint probabilities, a possible solution might be to use a numerical score as an alternative to the exact probability. A proper definition of such a score would greatly improve the performance of Bayesian Networks and at the same time would render a quasi-optimal solution.

Mixed-effects models

In this thesis we have shown the potential of Mixed-effects models in the context of multilevel epidemiology data. More precisely, we have satisfactorily applied Linear Mixed-effects Models (LMMs) to complex data comprising of both repeated and unrepeated individual measurements together with ecological or groupal variables.

Future research on the topic should consider the potential non-linear associations between exposures and outcomes using more complex correlation structures. A possible way to achieve this would be to use Nested Nonlinear Mixed-effects Models. Another possibility could be to apply Generalized Estimating Equations (GEEs), a method that allows to estimate the parameters of a Generalized Linear Model (GLM) with unknown correlation between outcomes.

From a more epidemiologic perspective, it will be important that future research further investigate the role of socioeconomic status to fully dismiss potential residual confounding by social class. This is a key point in this and other similar epidemiologic studies because both the outcome and exposures are usually highly correlated with the socioeconomic background of individuals. But, to success in this task, efforts must be directed at achieving a much better characterization of social class, possibly using many different proxy variables.

Differential expression

The study of differential expression has been revolutionised by the advent of Next Generation Sequencing technologies. RNA-seq offers an unprecented power to capture the real dynamics of gene-expression but, as the number of replicates grows, so does the underlying complexity of expression profiles. In this thesis we have presented a new method relying on the Poisson-Tweedie family of distributions for the analysis of differential expression using RNA-seq data.

Being a relatively young discipline in the world of science, RNA-seq is still in a development phase. As a result, considerable challenges and

issues arise that need to be effectively solved. Most of them are related to the experimental technologies needed for sequencing, library preparation, amplification, etc. But many others are more data related and would fall into the realm of the so called "Data Science". For instance, the limited amounts of material available per cell might lead to high levels of uncertainty about observations and appropriate statistic techniques would be needed to account for this variability. Sometimes amplification is used to generate more material, thus adding noise to the resulting data and this issue must also be taken into account in the analyses. Furthermore, the increase in resolution provided by the rapid development of the Next Generation Sequencing technologies results in multidimensional data, calling for scalable data analysis models and methods.

Regarding the Poisson-Tweedie family of distributions, we are of the opinion that the obtained results are promising and warrant further investigation. Future research should further develop its applications to the analysis of RNA-seq data. A natural progression would be to focus on the development of GLM models that accept the Poisson-Tweedie family as the distribution of the outcome variable. We have already devoted efforts to this task and have an effective initial model fitting algorithm although it still presents some issues when the parameter estimates get close to the boundaries of the parametric space. Once these issues have been circumvented, it would also be of interest to further develop the model into a Poisson-Tweedie Mixed-effects model, thus combining the covered methodologies of Papers 2, 3 and 4 of this thesis. We believe that the potential applications of this method in future epidemiology studies would be enormous.

Finally, in future research, more focus is needed to test the potential applications of the Poisson-Tweedie family of distributions to other research fields that heavily rely on count data such as econometrics, sociology or environmental sciences. We are aware that there have already been some attempts to apply the PT in several different contexts but, although some results are promising, they are still in a developmental phase and still require considerable efforts before they can be regarded as realistic alternatives to the current state of the art methods.

Appendix: other related published papers

I. Green spaces and cognitive development in primary schoolchildren

Green spaces and cognitive development in primary schoolchildren

Payam Dadvand[a,b,c,1], Mark J. Nieuwenhuijsen[a,b,c], Mikel Esnaola[a,b,c], Joan Forns[a,b,c,d], Xavier Basagaña[a,b,c], Mar Alvarez-Pedrerol[a,b,c], Ioar Rivas[a,b,c,e], Mónica López-Vicente[a,b,c], Montserrat De Castro Pascual[a,b,c], Jason Su[f], Michael Jerrett[g], Xavier Querol[e], and Jordi Sunyer[a,b,c,h]

[a]Centre for Research in Environmental Epidemiology (CREAL), 08003 Barcelona, Spain; [b]Experimental and Health Sciences, Pompeu Fabra University, 08003 Barcelona, Catalonia, Spain; [c]Ciber on Epidemiology and Public Health (CIBERESP), 28029 Madrid, Spain; [d]Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, 0473, Oslo, Norway; [e]Department of Geosciences, Institute of Environmental Assessment and Water Research, Spanish National Research Council (CSIC-IDEA), 08034 Barcelona, Catalonia, Spain; [f]Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA 94720-7360; [g]Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles, CA 90095; and [h]Hospital del Mar Medical Research Institute (IMIM), 08003 Barcelona, Catalonia, Spain

Edited by Susan Hanson, Clark University, Worcester, MA, and approved May 15, 2015 (received for review February 18, 2015)

Exposure to green space has been associated with better physical and mental health. Although this exposure could also influence cognitive development in children, available epidemiological evidence on such an impact is scarce. This study aimed to assess the association between exposure to green space and measures of cognitive development in primary schoolchildren. This study was based on 2,593 schoolchildren in the second to fourth grades (7–10 y) of 36 primary schools in Barcelona, Spain (2012–2013). Cognitive development was assessed as 12-mo change in developmental trajectory of working memory, superior working memory, and inattentiveness by using four repeated (every 3 mo) computerized cognitive tests for each outcome. We assessed exposure to green space by characterizing outdoor surrounding greenness at home and school and during commuting by using high-resolution (5 m × 5 m) satellite data on greenness (normalized difference vegetation index). Multilevel modeling was used to estimate the associations between green spaces and cognitive development. We observed an enhanced 12-mo progress in working memory and superior working memory and a greater 12-mo reduction in inattentiveness associated with greenness within and surrounding school boundaries and with total surrounding greenness index (including greenness surrounding home, commuting route, and school). Adding a traffic-related air pollutant (elemental carbon) to models explained 20–65% of our estimated associations between school greenness and 12-mo cognitive development. Our study showed a beneficial association between exposure to green space and cognitive development among schoolchildren that was partly mediated by reduction in exposure to air pollution.

neurodevelopment | greenness | cognition | built environment | school

Contact with nature is thought to play a crucial and irreplaceable role in brain development (1, 2). Natural environments including green spaces provide children with unique opportunities such as inciting engagement, risk taking, discovery, creativity, mastery and control, strengthening sense of self, inspiring basic emotional states including sense of wonder, and enhancing psychological restoration, which are suggested to influence positively different aspects of cognitive development (1–3). Beneficial effects of green spaces on cognitive development might accrue from direct influences such as those above, with green space itself exerting the positive influence or through indirect, mediated pathways. The ability of green spaces to mitigate traffic-related air pollution (TRAP) (4) could lead to a beneficial impact of green spaces on cognitive development, because exposure to TRAP has been negatively associated with cognitive development in children (5). Further to TRAP, green spaces can also reduce noise (6), which itself too has been negatively associated with cognitive development (7). Moreover, proximity to green spaces, particularly parks, has been suggested to increase physical activity (8), and higher levels of physical activity are related to improved cognitive development (9). Outdoor surrounding greenness has also been reported to enrich microbial input from the environment (10), which may positively influence cognitive development (10). Through these pathways, exposure to green space, including outdoor surrounding greenness and proximity to green spaces, could influence cognitive development in children, yet the available population-based evidence on the association between such exposure and cognitive development in children remains scarce.

The brain develops steadily during prenatal and early postnatal periods, which are considered as the most vulnerable windows for effects of environmental exposures (11). However, some cognitive functions closely related with learning and school achievement—such as working memory and attention—develop across childhood and adolescence as an essential part of cognitive maturation (12–14). We therefore hypothesized a priori that exposure to green space in primary schoolchildren could enhance cognitive development. Accordingly, our study aimed to assess the association between indicators of exposure to green space and measures of cognitive development, including working memory (the system that holds multiple pieces of transitory information in the mind where they can be manipulated), superior working memory (working memory that involves continuous updating of the working memory buffer), and inattentiveness in primary schoolchildren. As a secondary aim, we also evaluated the mediating role of a reduction in air pollution as one of the potential mechanisms underlying this association.

Significance

Green spaces have a range of health benefits, but little is known in relation to cognitive development in children. This study, based on comprehensive characterization of outdoor surrounding greenness (at home, school, and during commuting) and repeated computerized cognitive tests in schoolchildren, found an improvement in cognitive development associated with surrounding greenness, particularly with greenness at schools. This association was partly mediated by reductions in air pollution. Our findings provide policymakers with evidence for feasible and achievable targeted interventions such as improving green spaces at schools to attain improvements in mental capital at population level.

Author contributions: P.D., M.J.N., X.Q., and J. Sunyer designed research; M.J.N., J.F., M.A.-P., I.R., M.L.-V., M.D.C.P., X.Q., and J. Sunyer performed research; M.E., X.B., J. Su, and M.J. contributed new reagents/analytic tools; P.D., M.E., and X.B. analyzed data; and P.D. and J. Sunyer wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: pdadvand@creal.cat.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1503402112/-/DCSupplemental.

Methods

Study Setting. We undertook this study in Barcelona, Spain, a port city situated on the northeastern part of the Iberian Peninsula. It has a Mediterranean climate characterized by hot and dry summers, mild winters, and maximum precipitation and vegetation during autumn and spring. This study was conducted in the context of the brain development and air pollution ultrafine particles in school children (BREATHE) project. Of the 416 schools in Barcelona, 37 schools were initially selected to obtain maximum contrast in TRAP levels (i.e., nitrogen dioxide: NO_2), of which 36 accepted to participate and were included in the study (*SI Appendix*, Fig. S1). Participating schools were similar to the remaining schools in Barcelona in terms of the neighborhood socioeconomic vulnerability index (0.46 versus 0.50, Kruskal–Wallis test $P = 0.57$) and NO_2 levels (51.5 versus 50.9 $\mu g/m^3$, Kruskal–Wallis test $P = 0.72$).

All schoolchildren ($n = 4{,}562$) without special needs in the second to fourth grades (7–10 y) of these schools were invited to participate by letters or presentations in schools for parents, of which 2,623 (58%) agreed to take part in BREATHE. All children had been in the school for more than 6 mo (and 98% more than 1 y) before the beginning of the study. All parents or guardians signed the informed consent and the study was approved (No. 2010/41221/I) by the Clinical Research Ethical Committee of the Parc de Salut Mar, Barcelona.

Outcome: Cognitive Development. Cognitive development was assessed through 12-mo change in developmental trajectory of working memory and attention. We selected these functions because they grow steadily during preadolescence (15, 16). We used computerized *n*-back test for assessing working memory (15) and computerized attentional network test (ANT) (17) for evaluating attention.

From January 2012 to March 2013, children were evaluated every 3 mo over four repeated visits by using computerized tests in sessions lasting ∼40 min in length. Groups of 10–20 children wearing ear protectors were assessed together and supervised by one trained examiner per 3–4 children. For the *n*-back test, we examined different *n*-back loads (up to three-back) and stimuli (colors, numbers, letters, and words). For analysis here, we selected both two-back and three-back loads for number and word stimuli because they showed a clear age-dependent slope in the four measurements (4). The two-back predicts general mental abilities (i.e., working memory) whereas the three-back also predicts superior functions such as fluid intelligence (i.e., superior working memory) (18). All sets of *n*-back tests started with colors as a training phase to ensure participants' comprehension of the test. The *n*-back parameter analyzed was *d* prime (*d*′), a measure of detection subtracting the normalized false alarm rate from the hit rate [(Z hit rate − Z false alarm rate) ×100]. A higher *d*′ indicates more accurate test performance. Given that our final findings for numbers and words were similar, here we only show results for numbers. Among the ANT measures, we chose hit reaction time standard error (HRT-SE) (SE of RT for correct responses), a measure of response speed consistency throughout the test (19), because it showed a clear growth during the 1-y study period. A higher HRT-SE indicates highly variable reactions related to inattentiveness.

Exposure to Green Space. Our assessment of exposure to green space was based on a comprehensive characterization of outdoor surrounding greenness (photosynthetically active vegetation) encompassing greenness surrounding home, greenness surrounding commuting route between home and school (hereafter referred to as commuting greenness), and greenness within and around school boundaries.

To assess outdoor surrounding greenness we applied normalized difference vegetation index (NDVI) derived from RapidEye data at 5 m × 5 m resolution. NDVI is an indicator of greenness based on land surface reflectance of visible (red) and near-infrared parts of spectrum (20). It ranges between −1 and 1, with higher numbers indicating more greenness. The RapidEye Imagery is acquired from a constellation of five satellites 630 km above ground in sun-synchronous orbits. We generated our NDVI map by using the image obtained on July 23, 2012, that was available for our study region during our study period (*SI Appendix*, Fig. S1).

Residential surrounding greenness. Residential surrounding greenness was abstracted as the average of NDVI in a buffer of 250 m (21, 22) around the home address of each study participant. For 174 children (5.9%) who shared two homes, we used the address where the child spent most of her/his time.

Commuting greenness. Data on the main mode of commute to and from school was obtained from parents via questionnaires. Approximately 60% of participants reported walking as the main mode of commuting, whereas the 38% reported commuting by motor vehicles (private car, bus, motorcycle, or tram). The remaining 2% reported the underground metro train as the main mode of transport, for whom we assumed no exposure to greenness during commuting. For participants reporting walking as the main mode of commuting, we identified the shortest walking route to school and for participants reporting motor vehicles as the main mode of commuting, we identified the shortest driving route to school, based on street networks (network distance) by using network analyst extension from ArcGIS software v10. We defined commuting greenness as the average of NDVI in a 50-m buffer around the commuting route.

School greenness. To assess greenness within school premises, we first digitized the school boundaries and then averaged NDVI values within those boundaries. To assess greenness surrounding schools, we averaged NDVI values across a 50 m buffer around the school boundaries.

Total surrounding greenness index. We developed a total surrounding greenness index by averaging residential surrounding greenness (250-m buffer), commuting greenness, and greenness within school boundaries weighted by the daytime (12 h a day) that children were assumed to spend at home (3 h), commuting (1 h), and school (8 h). To avoid double-counting in developing this index, we abstracted as the average NDVI over commute corridor beyond the 250-m home buffer and 50-m school buffer.

Main Analyses. Data on 9,357 tests from 2,593 (99%) children were available for analysis. Because of the multilevel nature of the data (i.e., multiple visits for each child within schools), we used linear mixed effects models with the four repeated cognitive parameters as outcomes (one test at a time), each measure of exposure to green space (one at a time) as fixed effect predictor, and child and school as random effects (5). An interaction between age at each visit and the indicator of exposure to green space was included to capture changes in 12-mo progress in cognitive trajectory associated with greenness exposure (5). The main effect of exposure to green space, which was also included in the model, captured the baseline (visit 1) differences in cognitive function that were associated with exposure to green space before the first visit. This model was further adjusted for potential confounders identified a priori*:* age (centered at visit 1), sex, and indicators of socioeconomic status (SES) at both individual and area levels. Maternal education (no or primary/secondary/university) was used as the indicator of individual-level SES and Urban Vulnerability Index (23), a measure of neighborhood SES at the census tract (median area of 0.08 km^2 for the study region) was applied as the indicator of area-level SES. Linearity of the relation between exposure to green space and cognitive tests was assumed because generalized additive mixed models did not show any nonlinearity of associations. We estimated the change in average outcome scores associated with one interquartile range (IQR) increase (based on all study participants) in average NDVI. Statistical significance was set at $P < 0.05$. R statistical package was used to carry out the analyses.

Mediating Role of Traffic-Related Air Pollution. We hypothesized that reduction in TRAP levels could be one of the potential mechanisms underlying the association between greenness exposure and cognitive development. To quantify such a mediating role, we calculated the percent of the associations between greenness and cognitive development explained by TRAP as $[1 - (\beta_{gm}/\beta_g)] \times 100$, where β_{gm} was the regression coefficient for the greenness exposure in a fully adjusted model including the mediator (i.e., TRAP) and β_g was the regression coefficient in the fully adjusted model without including the mediator (24).

We focused on the associations between school greenness and cognitive development because they were the strongest among our evaluated associations (*Results*) and also because of the availability of data on levels of air pollutants at BREATHE schools that were monitored as part of the BREATHE project. Such a high-quality monitored data were not available for TRAP levels at homes or during commuting. Among the TRAPs monitored in the BREATHE framework, we chose indoor levels of elemental carbon (EC) for this mediation analyses. EC is mainly generated by fossil fuel combustion and is considered as a tracer of road traffic emissions in Barcelona (25). In other BREATHE analyses, we had observed that indoor EC was associated with adverse impacts on cognitive development (5) and EC levels were reduced in schools with higher greenness (4). Detailed description of TRAP sampling methodology at the BREATHE schools has been published (25, 26).

Results

Children were on average 8.5 y old at baseline and 50% were girls. Regarding maternal education, 13% of mothers had no or only primary school, 29% secondary school, and 58% university education. Further characteristics of the study participants are presented in *SI Appendix*, Table S1. Average working memory increased by 22.8%, superior working memory by 15.2%, and

inattentiveness decreased by 18.9% during the follow up (Table 1). At baseline, higher maternal education was associated with better cognitive function (*SI Appendix*, Table S2). For 12-mo progress, whereas higher maternal education was associated with larger reduction in inattentiveness, improvements in working memory and superior working memory were not associated with maternal education (*SI Appendix*, Table S2). The median (IQR) of our estimated surrounding greenness for all participants and across strata of maternal education are presented in Table 2 and *SI Appendix*, Table S2, respectively. The Spearman's correlation coefficient among residential, school, and commuting surrounding greenness varied from 0.46 (between surrounding greenness at home and greenness within school boundaries) to 0.80 (between commuting and school surrounding greenness) (*SI Appendix*, Table S3).

Main Analyses. We observed an enhanced 12-mo progress in working memory and superior working memory and a greater 12-mo reduction in inattentiveness associated with greenness within and surrounding school boundaries and with the total surrounding greenness index (Table 2, Fig. 1, and *SI Appendix*, Fig. S2). Commuting greenness was also associated with improved 12-mo progress in working memory and superior working memory, although the association for superior working memory was only marginally statistically significant. We did not observe any association between residential surrounding greenness and cognitive measurements (Table 2). None of the indicators of outdoor greenness were associated with baseline cognitive measurements (Table 2).

The findings for *n*-back tests with "word" stimuli were consistent with the aforementioned results for "number" stimuli (*SI Appendix*, Table S4). The association between commuting greenness and 12-mo progress in superior working memory, which had borderline statistical significance for the three-back test using number stimuli, was statistically significant for the test using word stimuli.

To explore the possibility of an impact of green space exposure on other ANT measures than inattentiveness, we repeated the main analyses by using alerting, orienting, and executive processing (one at a time) abstracted from ANT as outcome. We did not observe any statistically significant association for these outcomes with any of indicators of green space exposure (*SI Appendix*, Table S5), which was consistent with our observation that these measures did not show any clear growth during the study period.

We conducted a number of sensitivity analyses as described in *SI Appendix* that showed the robustness of our findings to alternative definition of total surrounding greenness index and commuting greenness and to including a range of relevant covariates in models (e.g., socioeconomic indicators and condition of venue at the time of cognitive tests).

Mediating Role of Traffic-Related Air Pollution. The Spearman's correlation coefficients between school EC levels and greenness within and surrounding school boundaries were −0.62 and −0.66 ($P < 0.01$), respectively. Adding EC to models explained 20–65% of associations between school greenness and 12-mo progress in cognitive functions (Table 3). Including EC reduced effect sizes in all models. EC made the associations between school surrounding greenness and superior working memory and between greenness within and surrounding school boundaries and inattentiveness much smaller and statistically nonsignificant (Table 3).

Discussion

To our knowledge, this is the first epidemiological study to report on the impact of exposure to green space on cognitive development in schoolchildren. School and total surrounding greenness index were associated with enhanced 12-mo progress in indicators of working memory and superior working memory and greater 12-mo reduction in inattentiveness. Commuting greenness was also associated with better 12-mo progress in working memory. Adding EC to our models explained 20–65% of our estimated associations between green spaces and 12-mo cognitive development.

Interpretation of Results. Over a 12-mo period, we observed that an IQR exposure increment in total surrounding greenness index was associated with a 5% increase in the progress of working memory, a 6% increase in the progress of the superior working memory, and a 1% reduction of inattentiveness. Among our assessed exposure measures, we observed the strongest associations for greenness within or surrounding school boundaries. Children spend a considerable part of their active daily time at schools and "green exercise" has been related to better mental health (27). Furthermore, the combination of physical activity in school with daily peaks of TRAPs in urban areas that often coincide with school time could result in a considerable inhaled dose of air pollutants at school. Consistently, in our other BREATHE analysis of the impact of TRAPs on cognitive development using the same measures of cognitive development as in this study, we also observed stronger associations for levels at school compared with those at home (5). Therefore, the ability of school greenness in reducing pollutant levels (4) might explain, in part, why we observed the strongest associations for school greenness.

We found some indications for an enhanced 12-mo progress in working memory associated with commuting greenness. Because of the strong correlation between greenness surrounding school boundaries and commuting greenness, it was not possible to determine the independent impact of commuting greenness (i.e., whether commuting greenness is a surrogate for school surrounding greenness). Therefore, our findings for commuting greenness should be interpreted with caution. To the best of our knowledge, this study is the first reporting on the potential impact of commuting greenness on health in general and on cognitive development in particular. We hypothesize that green exercise and visual access to greenness might underlie such an association, if any.

The beneficial associations for 12-mo progress in cognitive functions were stronger than those at baseline. Baseline estimates reflected the association between cognitive test scores at the first visit and the cumulative green space exposure preceding the study period, whereas our exposure assessment was based

Table 1. Description of the cognitive outcomes in children [median (25th–75th %)]

Visit	*n*	Age (mean), y	Working memory (WM) (two-back numbers), d′*	Superior WM (three-back numbers), d′*	Inattentiveness (ANT HRT-SE)†, ms
First visit	2,278	8.5	206 (129, 360)	112 (53, 171)	271 (205, 338)
Second visit	2,425	8.7	221 (129, 392)	112 (59, 190)	250 (186, 321)
Third visit	2,347	9.1	234 (129, 392)	128 (59, 190)	247 (183, 317)
Fourth visit	2,307	9.4	253 (152, 392)	129 (64, 210)	228 (165, 294)

*The *n*-back *d′* is a measure of detection subtracting the normalized false alarm rate from the hit rate [(Z hit rate − Z false alarm rate) × 100].
†Hit reaction time SE (HRT-SE), SE of reaction time for correct responses as a measure of response speed consistency throughout the test.

Table 2. Adjusted difference (95% confidence interval) in baseline and 12-mo progress of working memory, superior working memory, and inattentiveness per one interquartile range (IQR) change in greenness

		Working memory† (2-back number stimuli, d′)		Superior working memory† (3-back number stimuli, d′)		Inattentiveness† (HRT-SE, ms)	
Surrounding greenness	Median (IQR)	Baseline	Progress	Baseline	Progress	Baseline	Progress
Home	0.091 (0.053)	0.2 (-3.8, 4.2)	0.7 (-2.6, 4.1)	0.6 (-2.5, 3.7)	−0.1 (-2.7, 2.6)	2.0 (-1.4, 5.4)	−0.7 (-3.1, 1.7)
School							
Within	0.094 (0.085)	0.3 (−6.8, 7.4)	9.8 (5.2, 14.0)*	0.9 (−5.0, 6.8)	6.9 (3.4, 10.0)*	−4.0 (−12.0, 4.0)	−3.4 (−6.6, −0.2)*
Surrounding‡	0.100 (0.120)	3.2 (−4.3, 11)	9.5 (4.5, 15.0)*	1.5 (−4.8, 7.8)	6.3 (2.3, 10.0)*	−5.1 (−14.0, 3.6)	−3.7 (−7.3, −0.1)*
Commuting	0.100 (0.062)	1.5 (−3.5, 6.6)	4.9 (1.0, 8.8) *	3.5 (−0.6, 7.5)	3.1 (0.0, 6.1)	0.2 (−4.5, 4.9)	−1.2 (−4.0, 1.7)
Total surrounding greenness index	0.094 (0.073)	0.0 (−6.9, 6.5)	9.8 (5.0, 15.0)*	1.7 (−4.4, 7.8)	6.7 (2.8, 11.0)*	−2.4 (−9.8, 4.9)	−3.9 (−7.4, −0.4)*

*$P < 0.05$.
†Difference adjusted for age, sex, maternal education, and residential neighborhood socioeconomic status with school and subject as nested random effects.
‡Fifty-meter buffer around school boundaries.

on the home address of participants and the school they were attending during the study period, not including potential prior different addresses or schools to their current ones. Part of our observed larger estimates for 12-mo progress might therefore reflect better characterization of exposure, but it could also be due to the window of vulnerability for these high executive functions that develop significantly during the primary school age (12–14). This window of vulnerability might also explain why we observed the strongest associations for 12-mo progress in superior working memory that develops considerably during this period.

We did not observe any statistically significant difference in 12-mo progress in working memory and superior working memory (for which we found associations with green space exposure) between strata of maternal education. Moreover, further adjustment of our analyses for other indicators of SES like parental employment, marital status, and ethnicity (*SI Appendix, SI Methods*) did not change the interpretation of our findings notably. Furthermore, removing SES indicators (maternal education and neighborhood SES) from our fully adjusted models did not result in a considerable change in the interpretation of our findings (*SI Appendix*, Table S6). Additionally, we did not observe any statistically significant effect modification by maternal education or neighborhood SES for our associations ($P > 0.1$). These observations might suggest that our results were unlikely to have been affected by residual SES confounding.

Available Evidence and Potential Underlying Mechanisms. We are not aware of previous epidemiological studies on the impact of green space exposure on cognitive development in schoolchildren; therefore, it is not possible to compare our findings with those of others. Our findings, however, are consistent with several previous observations. Residential surrounding greenness has been related to better mental health including lower risk of depression and anxiety in children (28). Higher school greenness has been associated with better student performance at schools (29). Experimental studies have shown walking in nature or watching photos of nature could improve directed-attention abilities in adults (30) and have "therapeutic effects" on attention deficit hyperactivity disorder symptoms in children (31–34). Our previous cross-sectional analysis of BREATHE participants showed a protective impact of home and school greenness on behavioral problems including hyperactivity and inattention (35). That analysis was based on behavioral screening questionnaires rated by teachers and parents. In those questionnaires behavioral aspects that characterized hyperactivity/inattention were modestly correlated (Spearman's correlation coefficients ranging between 0.18 and 0.23) with the ANT inattentiveness score (at baseline) used in this study. A study by Wells (2000) reported that relocation to residences with higher "naturalness" improved cognitive function in a sample of 17 children (36). In an analysis of BREATHE schools, we observed that higher greenness inside and surrounding school boundaries was associated with lower TRAPs levels at schools (5), in line with our other study showing lower levels of personal exposure to TRAPs (based on personal monitors) associated with higher residential surrounding greenness in Barcelona (22). Another BREATHE analysis, using the same cognitive measures as the current study, demonstrated that higher levels of TRAPs at school were associated with diminished 12-mo cognitive progress (5). Thus, reduction of exposure to TRAPs associated with higher greenness could have partly underlain our observed associations. Consistently, in the current analysis we observed that including a TRAP (EC) in our models could explain one-fifth to two-thirds of the associations, suggesting that our observed beneficial associations between greenness exposure and cognitive development could have been partly mediated by reduction in exposure to TRAPs. These findings could also suggest that other mechanisms may account for 35–80% of our observed associations that was not explained by reduction in TRAP exposure. Higher ambient noise has been related with adverse impacts on cognitive development (7). The ability of green spaces to

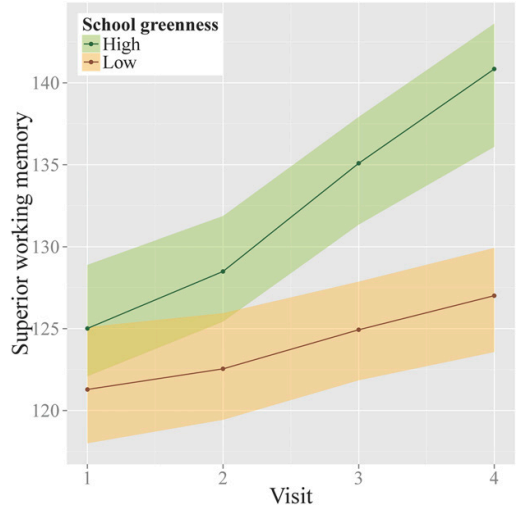

Fig. 1. Twelve-month progress (with 95% confidence bands) in superior working memory for participants with the first (low greenness) and third (high greenness) tertiles of greenness within the school boundaries.

Table 3. Difference (95% confidence interval) in 12-mo cognitive trajectory per one interquartile range change in greenness estimated by main analyses and models further including school indoor elemental carbon (EC) interaction with age

Outcomes/exposures	Main analyses[†,‡]	Further adjusted for EC[‡]	% explained
Working memory			
Within school	9.8 (5.2, 14.0)*	8.7 (2.5, 15.0)*	20.4
Surrounding school	9.5 (4.5, 15.0)*	6.9 (0.9, 13.0)*	27.4
Superior working memory			
Within school	6.9 (3.4, 10.0)*	4.9 (0.1, 9.8)*	29.0
Surrounding school[§]	6.3 (2.3, 10.0)*	3.3 (-1.5, 8.1)	47.6
Inattentiveness			
Within school	−3.4 (-6.6, -0.2)*	−1.2 (-5.6, 3.2)	64.7
Surrounding school	−3.7 (-7.3, -0.1)*	−1.8 (-6.1, 2.5)	51.4

*$P < 0.05$.
[†]Adjusted for age, sex, maternal education, and residential neighborhood socioeconomic status with school and subject as nested random effects.
[‡]Estimates per 0.085 and 0.120 change respectively in greenness within and surrounding school boundaries (i.e., 1-interquartile change).
[§]Fifty-meter buffer around school boundaries.

reduce noise (6) might therefore explain a part of our observed associations (37). Moreover, proximity to green spaces has been reported to increase physical activity (38), and physical activity has been associated with better cognitive function in children (9). Furthermore, parental psychological stress and depression have been reported to be adversely associated with cognitive development in their children (39) and exposure to green space has been associated with evidence of stress restorative effects and reduced depression in adults (3, 28). A growing body of evidence also suggests that a failure of the immunoregulatory pathways due to a reduced exposure to macroorganisms and microorganisms in Westernized populations might play a role in impairment of brain development (10, 40) with childhood as a particular window of vulnerability (41). Therefore, the ability of outdoor surrounding greenness to enhance immunoregulation-inducing microbial input from the environment (10) could have been another mechanism underlying our observed association between greenness exposure and cognitive development.

Implications for Policymakers. Approximately one-half of the world population lives in cities, and it is projected that by 2030, three of every five persons will live in urban areas worldwide (42). Urban areas are characterized by a network of nonnatural built-up infrastructures with increased pollutant levels and less green environments (43). Children's exposure to these pollutants such as air pollution and noise has been associated with detrimental impacts on their cognitive development. Our findings suggest for a beneficial impact of green space exposure on cognitive development, with part of this effect resulting from buffering against such urban environmental pollutants. This impact was more evident for surrounding greenness at school and for working memory and superior working memory, which are predictors of learning and academic attainment (44). Schoolchildren with a superior working memory progress of less than one-10th of a percentile (45) of the distribution can be classified as impaired superior working memory progress. Our results suggest that if schools increased greenness within their boundaries by the observed IQR (Fig. 1), then 8.8% of children with impaired superior working memory progress would move out of this category. Our findings, therefore, hold importance for policymakers when translating evidence into feasible and achievable targeted interventions such as improving greenness at schools, given that improved cognitive development in children attending schools with more greenness could result in an advantage in mental capital, which, in turn, would have lasting effects through the life-course.

Strengths and Limitations of Study. This study was based on repeated computerized tests of cognitive development to quantify different aspects of cognitive development in study participants. These tests have been reported to have acceptable internal consistency, reasonable factorial structure, and good criterion validity and statistical dependencies for use in general population (46). We applied one of the most comprehensive approaches to date to assess exposure to green space by characterizing the outdoor surrounding greenness at home and school and during commuting by using high-resolution (5 m × 5 m) satellite data on greenness, enabling us to account for small-area green spaces (e.g., home gardens, street trees, and green verges) in a standardized way.

Our study also faced some limitations. The generalizability of our findings might have been affected by selection bias in that those participants participated in BREATHE were different from those not participated with respect to SES. Approximately 58% of mothers in our study population had a university degree, which was higher that the regional average of 50% among women between 25 and 39 y old living in Barcelona (47). We did not, however, observe any indication of effect modification by maternal education in our associations. Moreover, the Urban Vulnerability Index of the schools was not associated with school participation rate (Spearman's correlation coefficient = −0.09, P = 0.61); these observations might suggest that the socioeconomic status was less likely to be a major predictor of participating in the study. Similarly, school greenness was not associated with participation rate at schools (Spearman's correlation coefficients of −0.06 with P value = 0.72 for greenness within school boundaries and 0.13 with P value = 0.43 for greenness surrounding schools). Our exposure assessment focused on exposure during the school age, overlooking other potential windows of susceptibility such as prenatal and preschool periods. Investigating these windows of susceptibility presents an opportunity for future studies. By using an NDVI map obtained at a single point in time (2012), we effectively assumed that the spatial distribution of NDVI across our study region remained constant over the study period (2012). The findings of our previous studies support the stability of the NDVI spatial contrast over seasons and years (21, 48). Finally, data were not available for some potentially relevant confounders, such as parental mental health status.

Conclusions

Exposure to outdoor surrounding greenness was associated with a beneficial impact on cognitive development in schoolchildren.

These associations were only partly mediated by reduction in TRAP levels, suggesting that other mechanisms likely underlie this association. Our observed beneficial associations were consistent for working memory, superior working memory, and inattentiveness and were more evident for greenness at school. Further studies are warranted to replicate our findings in other settings with different climates and to investigate other cognitive functions with different windows of susceptibility such as prenatal and preschool periods.

ACKNOWLEDGMENTS. We thank all the families and schools participating in the study for their altruism and their collaboration; Xavier Mayoral for the technical support of the *n*-back test; and Cecilia Persavento, Judit Gonzalez, Laura Bouso, and Pere Figueras for conducting the field work. The research leading to these results has received funding from the European Research Council (ERC) under ERC Grant Agreement 268479—the BREATHE project. The research (PHENOTYPE) leading to the methodology applied for the exposure assessment in this study has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under Grant Agreement 282996. P.D. is funded by Ramón y Cajal Fellowship RYC-2012-10995 awarded by the Spanish Ministry of Economy and Competitiveness.

1. Kahn PH, Kellert SR (2002) *Children and Nature: Psychological, Sociocultural, and Evolutionary Investigations* (MIT Press, Cambridge, MA).
2. Kellert SR (2005) *Building for Life: Designing and Understanding the Human-Nature Connection* (Island, Washington).
3. Bowler DE, Buyung-Ali LM, Knight TM, Pullin AS (2010) A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health* 10:456.
4. Dadvand P, et al. (2015) The association between greenness and traffic-related air pollution at schools. *Sci Total Environ* 523:59–63.
5. Sunyer J, et al. (2015) Association between traffic-related air pollution in schools and cognitive development in primary school children: A prospective cohort study. *PLoS Med* 12(3):e1001792.
6. Gidlöf-Gunnarsson A, Öhrström E (2007) Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. *Landsc Urban Plan* 83(2):115–126.
7. Klatte M, Bergström K, Lachmann T (2013) Does noise affect learning? A short review on noise effects on cognitive performance in children. *Front Psychol* 4:578.
8. James P, Banay RF, Hart JE, Laden F (2015) A review of the health benefits of greenness. *Curr Epidemiol Reports* 2(2):131–142.
9. Fedewa AL, Ahn S (2011) The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: A meta-analysis. *Res Q Exerc Sport* 82(3):521–535.
10. Rook GA (2013) Regulation of the immune system by biodiversity from the natural environment: An ecosystem service essential to health. *Proc Natl Acad Sci USA* 110(46):18360–18367.
11. Grandjean P, Landrigan PJ (2014) Neurobehavioural effects of developmental toxicity. *Lancet Neurol* 13(3):330–338.
12. Anderson P (2002) Assessment and development of executive function (EF) during childhood. *Child Neuropsychol* 8(2):71–82.
13. Ullman H, Almeida R, Klingberg T (2014) Structural maturation and brain activity predict future working memory capacity during childhood development. *J Neurosci* 34(5):1592–1598.
14. Østby Y, Tamnes CK, Fjell AM, Walhovd KB (2011) Morphometry and connectivity of the fronto-parietal verbal working memory network in development. *Neuropsychologia* 49(14):3854–3862.
15. Jaeggi SM, Buschkuehl M, Perrig WJ, Meier B (2010) The concurrent validity of the N-back task as a working memory measure. *Memory* 18(4):394–412.
16. Rueda MR, Rothbart MK, McCandliss BD, Saccomanno L, Posner MI (2005) Training, maturation, and genetic influences on the development of executive attention. *Proc Natl Acad Sci USA* 102(41):14931–14936.
17. Rueda MR, et al. (2004) Development of attentional networks in childhood. *Neuropsychologia* 42(8):1029–1040.
18. Shelton JT, Elliott EM, Matthews RA, Hill BD, Gouvier WD (2010) The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *J Exp Psychol Learn Mem Cogn* 36(3):813–820.
19. Conners CK, Staff MHS (2000) *Conners' Continuous Performance Test II: Computer Program for Windows Technical Guide and Software Manual* (Mutli-Health Systems, North Tonwanda, NY).
20. Weier J, Herring D (2011) *Measuring Vegetation (NDVI & EVI)* (Natl Aeronaut Space Admin, Greenbelt, MD).
21. Dadvand P, et al. (2012) Surrounding greenness and pregnancy outcomes in four Spanish birth cohorts. *Environ Health Perspect* 120(10):1481–1487.
22. Dadvand P, et al. (2012) Surrounding greenness and exposure to air pollution during pregnancy: An analysis of personal monitoring data. *Environ Health Perspect* 120(9): 1286–1290.
23. Spanish Ministry of Public Works (2012) *Atlas of Urban Vulnerability in Spain. Methodology and Contents*, ed Aja AH (Spanish Ministry of Public Works, Madrid).
24. Preacher KJ, Kelley K (2011) Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychol Methods* 16(2):93–115.
25. Amato F, et al. (2014) Sources of indoor and outdoor PM2.5 concentrations in primary schools. *Sci Total Environ* 490:757–765.
26. Rivas I, et al. (2014) Child exposure to indoor and outdoor air pollutants in schools in Barcelona, Spain. *Environ Int* 69:200–212.
27. Thompson Coon J, et al. (2011) Does participating in physical activity in outdoor natural environments have a greater effect on physical and mental wellbeing than physical activity indoors? A systematic review. *Environ Sci Technol* 45(5):1761–1772.
28. Maas J, et al. (2009) Morbidity is related to a green living environment. *J Epidemiol Community Health* 63(12):967–973.
29. Wu C-D, et al. (2014) Linking student performance in Massachusetts elementary schools with the "greenness" of school surroundings using remote sensing. *PLoS ONE* 9(10):e108548.
30. Berman MG, Jonides J, Kaplan S (2008) The cognitive benefits of interacting with nature. *Psychol Sci* 19(12):1207–1212.
31. van den Berg AE, van den Berg CG (2011) A comparison of children with ADHD in a natural and built setting. *Child Care Health Dev* 37(3):430–439.
32. Taylor AF, Kuo FE (2009) Children with attention deficits concentrate better after walk in the park. *J Atten Disord* 12(5):402–409.
33. Taylor AF, Kuo FE, Sullivan WC (2001) Coping with ADD: The surprising connection to green play settings. *Environ Behav* 33(1):54–77.
34. Kuo FE, Taylor AF (2004) A potential natural treatment for attention-deficit/hyperactivity disorder: Evidence from a national study. *Am J Public Health* 94(9):1580–1586.
35. Amoly E, et al. (2014) Green and blue spaces and behavioral development in Barcelona schoolchildren: The BREATHE project. *Environ Health Perspect* 122(12): 1351–1358.
36. Wells NM (2000) At home with nature effects of greenness on children's cognitive functioning. *Environ Behav* 32(6):775–795.
37. Stansfeld SA, et al.; RANCH study team (2005) Aircraft and road traffic noise and children's cognition and health: A cross-national study. *Lancet* 365(9475):1942–1949.
38. Lee AC, Maheswaran R (2011) The health benefits of urban green spaces: A review of the evidence. *J Public Health (Oxf)* 33(2):212–222.
39. Ramchandani P, Psychogiou L (2009) Paternal psychiatric disorders and children's psychosocial development. *Lancet* 374(9690):646–653.
40. Rook GAW, Lowry CA, Raison CL (2013) Microbial 'Old Friends', immunoregulation and stress resilience. *Evol Med Public Health* 2013(1):46–64.
41. Rook GAW, Lowry CA, Raison CL (April 13, 2014) Hygiene and other early childhood influences on the subsequent function of the immune system. *Brain Res*, 10.1016/j.brainres.2014.04.004.
42. Martine G, Marshall A (2007) *State of World Population 2007: Unleashing the Potential of Urban Growth* (United Nations Popul Fund, New York).
43. Escobedo FJ, Kroeger T, Wagner JE (2011) Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environ Pollut* 159(8-9):2078–2087.
44. Alloway TP, Alloway RG (2010) Investigating the predictive roles of working memory and IQ in academic attainment. *J Exp Child Psychol* 106(1):20–29.
45. Lezak MD, Howieson DB, Loring DW, Hannay HJ, Fischer JS (2004) *Neuropsychological Assessment* (Oxford Univ Press, New York).
46. Forns J, et al. (2014) The n-back test and the attentional network task as measures of child neuropsychological development in epidemiological studies. *Neuropsychology* 28(4):519–529.
47. Barcelona City Council (2013) *Statistical Yearbook of Barcelona City. Year 2013.* (Barcelona City Council, Barcelona).
48. Dadvand P, et al. (2014) Inequality, green spaces, and pregnant women: Roles of ethnicity and individual and neighbourhood socioeconomic status. *Environ Int* 71: 101–108.

II.Exposure to road traffic noise and cognitive development in schoolchildren in Barcelona, Spain: A population-based cohort study

PLOS MEDICINE

RESEARCH ARTICLE

Exposure to road traffic noise and cognitive development in schoolchildren in Barcelona, Spain: A population-based cohort study

Maria Foraster[1,2,3,4]*, Mikel Esnaola[1,2,3], Mónica López-Vicente[5], Ioar Rivas[1,2,3], Mar Álvarez-Pedrerol[1,2,3], Cecilia Persavento[1,2,3], Nuria Sebastian-Galles[2], Jesus Pujol[6,7], Payam Dadvand[1,2,3], Jordi Sunyer[1,2,3,8]

1 ISGlobal, Barcelona, Spain, **2** Universitat Pompeu Fabra (UPF), Barcelona, Spain, **3** CIBER Epidemiología y Salud Pública (CIBERESp), Madrid, Spain, **4** PHAGEX Research Group, Blanquerna School of Health Science, Universitat Ramon Llull (URL), Barcelona, Spain, **5** Department of Child and Adolescent Psychiatry and Psychology, Erasmus MC University Medical Center, Rotterdam, the Netherlands, **6** MRI Research Unit, Department of Radiology, Hospital del Mar, Barcelona, Spain, **7** Centro Investigación Biomédica en Red de Salud Mental, CIBERSAM, Barcelona, Spain, **8** IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

* maria.foraster@isglobal.org

OPEN ACCESS

Citation: Foraster M, Esnaola M, López-Vicente M, Rivas I, Álvarez-Pedrerol M, Persavento C, et al. (2022) Exposure to road traffic noise and cognitive development in schoolchildren in Barcelona, Spain: A population-based cohort study. PLoS Med 19(6): e1004001. https://doi.org/10.1371/journal.pmed.1004001

Academic Editor: Stephen Stansfeld, Wolfson Institute of Preventive Medicine, UNITED KINGDOM

Received: October 13, 2021

Accepted: April 27, 2022

Published: June 2, 2022

Data Availability Statement: Data cannot be shared publicly because they contain personal information of children and according to the ethical approval they should be kept confidential. Data are available from the Institutional Data Access / Ethics officer (contact via: research.management@isglobal.org, https://www.isglobal.org/en/-/breathe-brain-development-and-air-pollution-ultrafine-particles-in-school-children) for

Abstract

Background

Road traffic noise is a prevalent and known health hazard. However, little is known yet about its effect on children's cognition. We aimed to study the association between exposure to road traffic noise and the development of working memory and attention in primary school children, considering school-outdoor and school-indoor annual average noise levels and noise fluctuation characteristics, as well as home-outdoor noise exposure.

Methods and findings

We followed up a population-based sample of 2,680 children aged 7 to 10 years from 38 schools in Barcelona (Catalonia, Spain) between January 2012 to March 2013. Children underwent computerised cognitive tests 4 times ($n = 10,112$), for working memory (2-back task, detectability), complex working memory (3-back task, detectability), and inattentiveness (Attention Network Task, hit reaction time standard error, in milliseconds). Road traffic noise was measured indoors and outdoors at schools, at the start of the school year, using standard protocols to obtain A-weighted equivalent sound pressure levels, i.e., annual average levels scaled to human hearing, for the daytime (daytime LAeq, in dB). We also derived fluctuation indicators out of the measurements (noise intermittency ratio, %; and number of noise events) and obtained individual estimated indoor noise levels (LAeq) correcting for classroom orientation and classroom change between years. Home-outdoor noise exposure at home (Lden, i.e., EU indicator for the 24-hour annual average levels) was estimated using Barcelona's noise map for year 2012, according to the European Noise Directive (2002). We used linear mixed models to evaluate the association between exposure to noise and cognitive development adjusting for age, sex, maternal education,

researchers who meet the criteria for access to confidential data.

Funding: This research received funding from the following sources: European Research Council under the ERC Grant Agreement number 268479 (FP7 Ideas) – BREATHE project (JS) and the ERC Grant agreement number 785994 (H2020) - AirNB project (JS), https://erc.europa.eu/. Project "PI16/00118", funded by Instituto de Salud Carlos III and co-funded by European Union (European Regional Development Fund) "A way to make Europe" (JS), www.isciii.es. Agència de Gestió d'Ajuts Universitaris (Beatriu de Pinós fellowship, 2017 BP 00173, MF), https://agaur.gencat.cat/en/Beatriu-de-Pinos. MF is currently supported by an AXA Research Fund grant, https://www.axa-research.org/. We also acknowledge support from the Spanish Ministry of Science and Innovation and State Research Agency through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S, https://www.ciencia.gob.es/site-web/en/Organismos-y-Centros/Centros-y-Unidades-de-Excelencia.html, and the Generalitat de Catalunya through the CERCA Program (https://cerca.cat/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ANT, Attentional Network Test; BREATHE, Brain Development and Air Pollution Ultrafine Particles in School Children; *d'*, detectability; EC, elemental carbon; HRT-SE, hit reaction time standard error; IR, Intermittency Ratio; IQR, interquartile range; LAeq, A-weighted equivalent noise levels; Lday, LAeq for the daytime (7 AM to 9 PM); Lden, LAeq for the 24 hours with 5 dB and 10 dB penalties for the evening (9 PM to 11 PM) and nighttime (11 PM to 7 AM), respectively; NE, noise event; NO_2, nitrogen dioxide; PISA, Programme for International Student Assessment; RANCH, road traffic and aircraft noise exposure and children's cognition and health; SDQ, Strengths and Difficulties Questionnaire; TRAP, traffic-related air pollution.

socioeconomical vulnerability index at home, indoor or outdoor traffic-related air pollution (TRAP) for corresponding school models or outdoor nitrogen dioxide (NO_2) for home models. Child and school were included as nested random effects.

The median age (percentile 25, percentile 75) of children in visit 1 was 8.5 (7.8; 9.3) years, 49.9% were girls, and 50% of the schools were public. School-outdoor exposure to road traffic noise was associated with a slower development in working memory (2-back and 3-back) and greater inattentiveness over 1 year in children, both for the average noise level (e.g., −4.83 points [95% CI: −7.21, −2.45], *p*-value < 0.001, in 2-back detectability per 5 dB in street levels) and noise fluctuation (e.g., −4.38 [−7.08, −1.67], *p*-value = 0.002, per 50 noise events at street level). Individual exposure to the road traffic average noise level in classrooms was only associated with inattentiveness (2.49 ms [0, 4.81], *p*-value = 0.050, per 5 dB), whereas indoor noise fluctuation was consistently associated with all outcomes. Home-outdoor noise exposure was not associated with the outcomes. Study limitations include a potential lack of generalizability (58% of mothers with university degree in our study versus 50% in the region) and the lack of past noise exposure assessment.

Conclusions

We observed that exposure to road traffic noise at school, but not at home, was associated with slower development of working memory, complex working memory, and attention in schoolchildren over 1 year. Associations with noise fluctuation indicators were more evident than with average noise levels in classrooms.

Author summary

Why was this study done?

- Exposure to aircraft noise has been associated with impaired cognitive development in schoolchildren, and experiments have also observed that animals exposed to moderate or high noise levels for 4 to 30 days suffer changes in the brain.
- Road traffic noise is the most common noise source and many children are exposed to it at school; however, it is still unclear whether it affects children's cognitive development, including important aspects such as working memory or attention.
- Moreover, while previous studies have assessed exposure to the average noise level outside the school, none have assessed whether the noise peaks of fluctuating traffic and exposure to noise inside the classroom could affect children's cognition.

What did the researchers do and find?

- We carried out a cohort study to assess whether school exposure (inside and outside classrooms) and home exposure to road traffic noise were associated with the development of working memory and attention over 12 months in 2,680 children aged 7 to 10 years from 38 schools in Barcelona, Spain.

- We assessed long-term exposure to road traffic noise outside and inside the school with measurements (average noise levels and noise fluctuation) and at home with Barcelona's noise map (outdoor average noise levels).
- We used computerised tests every 3 months over 12 months to measure the development of working memory (the system that keeps and manipulates transitory information), complex working memory (it further involves continuous updating of the working memory), and inattentiveness in children.
- We observed that higher exposure to road traffic noise outside and inside the school, but not at home, was associated with a slower development of working memory and a slower improvement of inattentiveness over 12 months. Inside the classroom, associations were more evident for exposure to noise fluctuation than to average noise levels.

What do these findings mean?

- These findings suggest that, in children aged 7 to 10 years in Barcelona, higher exposure to road traffic noise at school relates to poorer development of attention and working memory. These are important for learning.
- The findings might not be applicable to other populations and need replication in other locations.
- These findings are of public health relevance given the large number of children exposed to road traffic noise in schools and support the implementation of environmental noise policies that protect the school environment.

Introduction

Road traffic noise is the most prevalent environmental and transportation noise source in Europe [1]. Transportation noise is the second most detrimental environmental factor for ill health in Europe, just after air pollution [2,3]. Such health impact is supported by an increasing number of epidemiological studies in adults; however, little is known yet about the effects in children [4–6].

One of the first adverse effects of noise on children could relate to cognitive development, given that childhood is a vulnerable period for brain maturation [7]. It is suggested that noise may impact cognitive abilities directly or lead to impaired attention, frustration, learned helplessness, arousal, or tuning out, which could impact performance and learning in the long term [8]. The impact of noise on cognition is supported by animal experiments in rats in which subchronic exposure to white noise (4 hours/day, 100 dB, up to 30 days) led to electroencephalographic changes in the occipital and prefrontal regions (relevant for executive functions) [9,10] and to reduced dendrite number in the hippocampus (relevant for learning and memory) [10]. An experiment exposing mice to average noise levels closer to environmental noise levels (72 dB continuous aircraft noise, up to 4 days) further observed increased oxidative stress in the brain, particularly in the frontal cortex [11].

In schoolchildren, there is substantial evidence for the association between exposure to aircraft noise and decreased cognitive development, particularly for reading comprehension, achievement tests, and long-term memory, according to longitudinal studies, intervention studies, and recent systematic reviews [5,8]. However, the available evidence for the association between exposure to road traffic noise and children's cognition is still limited and based only on cross-sectional studies [5,8]. In specific, it is unclear how noise could affect working memory or attention, which are essential for learning and school attainment [12,13] and which develop actively between 6 to 10 years of age as a result of cognitive maturation [14–16]. To our knowledge, only 5 cross-sectional studies from 3 different projects have examined the association between exposure to road traffic noise and tests of attention and they used different tests yielding mixed results [17–21]. Another 5, all part of the same cross-sectional RANCH study (road traffic and aircraft noise exposure and children's cognition and health), evaluated the association between exposure to road traffic noise and tests of working memory and reported no effects [19–23]. All of these studies except Cohen and colleagues [17], evaluated school-outdoor road traffic noise levels, using measured or modelled A-weighted equivalent noise levels (LAeq), which were representative of annual average noise levels. Cohen and colleagues [17] only assessed home-outdoor road traffic noise exposure, but using floor level as a proxy of noise exposure. Van Kempen and colleagues [20,21] assessed both school-outdoor and home-outdoor average levels of road traffic noise, and only observed associations between school-outdoor noise exposure and worse attention.

Importantly, none of the few previous studies have evaluated indoor road traffic noise levels in the classroom, where repeated exposure to noise could affect concentration, learning, and cognition [7]. Moreover, studies have focused on exposure to average noise levels, whereas the role of noise fluctuation (i.e., the presence of peaks and noise intermittency) on children's cognition remains unknown. Indirect evidence in adults indicates an association of noise fluctuation with endothelial dysfunction [24] and higher annoyance in people exposed to greater road traffic noise intermittency [25]. Finally, only 2 RANCH studies adjusted associations for coexposure to traffic-related air pollution [21,22], which has been also related with cognitive impairment in children [26]. In conclusion, there is a need for longitudinal studies and of novel, comprehensive assessments of children's exposure to noise at school and home in order to understand the effects of exposure to road traffic noise on children's cognitive development.

We hypothesise that school and home exposure to road traffic noise impairs the development of working memory and attention in children. The aim of the present study was to assess the association between exposure to road traffic noise indoors and outdoors at schools, accounting both for the annual average noise level and the noise fluctuation characteristics, with the development of working memory and inattentiveness over 12 months in primary schoolchildren, based on the BREATHE project (Brain Development and Air Pollution Ultrafine Particles in School Children), a cohort study of children in Barcelona. The study also assessed the association between home-outdoor long-term exposure to road traffic noise with the development of working memory and inattentiveness.

Methodology

Study population and design

This longitudinal study was carried out in the city of Barcelona, within the context of the BREATHE project (2011 to 2016). The BREATHE project aimed to study how air pollutants, including noise pollution, affected children's cognition. Barcelona is a densely populated city in northeast of Spain, in which the main source of noise is road traffic noise. In Barcelona, a total of 53% of the population are exposed to road traffic noise levels above the recommended

levels by the World Health Organization (24-hour EU noise indicator: Lden $\geq$ 53 dB) [4,27]. The BREATHE project assessed 2,897 children aged 7 to 10 years who were followed up through 2 academic courses between years 2012 and 2013 with a total of 4 visits and who attended 39 schools in Barcelona (Catalonia, Spain). Schools were selected based on the range of estimated outdoor levels of traffic-related nitrogen dioxide (NO_2) [28] in Barcelona, which would therefore also represent the range of noise as another traffic-related factor. We paired schools with low and high NO_2 levels by socioeconomic vulnerability index (SES) and type of school (i.e., private/public). In specific, we identified schools in Barcelona with low to moderate NO_2 levels (maximum 51 $\mu g/m^3$, mean of the city) estimated with a land use regression model [28] and searched schools with higher NO_2 levels but of similar type and SES (S1 Table). The participating schools were representative of the rest of schools in Barcelona in terms of neighbourhood socioeconomic vulnerability index (0.46 versus 0.50, Kruskal–Wallis test, p-value = 0.570) and NO_2 levels (51.5 versus 50.9 $\mu g/m^3$, Kruskal–Wallis test, p-value = 0.720). We invited all children from second to fourth grade who had no special needs and a 59% of families agreed to participate. All children had attended the same school for at least 6 months and a 98% for more than 1 year. In the present study, we excluded 1 school due to incomplete noise data. This study did not have a prespecified analysis plan.

All the participating children's parents or guardians signed and provided the informed consent. The research procedures were explained in detail to the children, who could ask any question. All included children provided their assent. The study received ethical approval by the Clinical Research Ethical Committee (No. 2010/41221/I) of the Institut Hospital del Mar d'Investigacions Mèdiques–Parc de Salut Mar, Barcelona, Spain, and the FP7-ERC-2010-AdG Ethics Review Committee.

This study is reported as per the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline (S1 STROBE Checklist).

Noise exposure assessment

Measured average noise levels at school. Environmental campaigns started in January 2012 and were performed twice, 6 months apart, during the cold and hot seasons. School pairs were assessed simultaneously. Road traffic noise was measured indoors (in classrooms) in the first campaign and indoors/outdoors 6 months later during the second campaign, following ISO 1996–2 (2007) for long-term environmental noise assessment based on short-term measurements. As part of this protocol, we carried out supervised time-stamped 30-min recordings of A-weighted equivalent noise levels (LAeq, in dB) during 2 consecutive weekdays before lessons started, using a calibrated CESVA SC160 device (type II microphone). We took simultaneous measurements indoors in 1 classroom (LAeq,in) and outdoors in the playground (LAeq,playgr.), followed right after by a street recording in front of the school (LAeq,street). Windows were closed during the measurement. Noise artefacts were subtracted to obtain clean road noise measurements. The 2-day mean of the measurements at each location during the second campaign, which assessed all school environments, was used as the exposure estimate.

Postprocessing analysis further confirmed the validity of the short-term noise protocol (ISO 1996–2) and the use of the 2-day mean to represent long-term (i.e., annual) means. Indoor noise measurements exhibited high reproducibility, i.e., intraclass correlation between the 2-day mean of each campaign = 0.89 and of the mean of the 2 campaigns = 0.94 (p-values < 0.001, 1-way random effects model). Moreover, the LAeq, street average was highly correlated (Spearman's rank r = 0.86, Pearsons' r = 0.85, p-values < 0.001) with the annual average noise levels for the daytime (Lday) obtained at the same location from the 2012

Strategic Noise Map for Barcelona derived under the EU Directive 2002/49/EC [29]. The noise map was published in year 2012 and represents average noise levels before year 2012.

To provide individual indoor average noise levels for each child during the study period (LAeq,in-indiv.), we further modelled indoor average levels of road traffic noise for all classrooms (adjusted R^2 = 67.7%; see S1 Text), based on data on the floor level of classrooms, room orientation (classroom oriented towards: indoor area, outdoor courtyard, or directly to the street), outdoor levels of road traffic noise, the type of courtyard (open, semi-open, built), and the type of windows, among others, and accounted for child's change of classroom between school years.

Measured noise fluctuation at school. We calculated the average number of individual noise events (NE) at each of the measured locations, i.e., street (NE, street), playground (NE, playgr.), and indoors (NE, in) as defined by [30] and previously used [24]. A noise event was labelled as a noise peak whose maximum noise level exceeded 3 dB above the total LAeq level during the measurement period [30]. Subsequently, noise intermittency ratio at the measured locations (IR, street; IR, playgr.; and IR, in) was calculated as the ratio between the LAeq noise level of the noise events for period T (Leq,T,events) and the total LAeq noise level for the same period T (Leq,T,tot), both expressed in sound energy (unit: percent) (Eq 1).

$$IR = \frac{10^{0.1\mathrm{Leq,T,events}}}{10^{0.1\mathrm{Leq,T,tot}}} x\ 100 \quad (1)$$

These fluctuation metrics were designed by Wunderli and colleagues [30], after the original planning of the current study. These metrics were calculated from the original noise measurement data.

Modelled average noise levels at school and home. We assigned exposure to road traffic noise as annual average road traffic noise levels at the geocoded noise measurement location in front of the school and at the geocoded home address of each participant, using the 2012 Strategic Noise Map of Barcelona, derived under the European Directive 2002/EC/49 [29]. Assignment of modelled noise levels was only possible in the 34 schools and 2,346 home addresses that fell within the boundaries of the city noise map, which excludes the outskirts of the city where 4 schools and some children's homes were located. Specifically, at schools, we estimated the A-weighted equivalent noise levels for the daytime (Lday, from 7 AM to 9 PM) to account for children's exposure during the academic time window. At home, we estimated the standard EU indicator for the 24 hours, i.e., the A-weighted equivalent noise levels for the 24 hours (Lden) with 5 dB and 10 dB penalties for the evening (9 PM to 11 PM) and nighttime (11 PM to 7 AM) to account for the children's exposure during the nonacademic hours, weekends, and holidays during the year.

Outcomes

From January 2012 to March 2013, children were visited 4 times (every 3 months) to carry out computerised psychometric measurements for working memory and inattentiveness in sessions of approximately 40 minutes. Tests were performed at schools by trained fieldworkers, who noted any incidents (including noise in the room) during the test. All children wore earphones to perform the tasks, which limited any influence of external noise stimuli on their outcomes.

We assessed 2 cognitive functions that develop rapidly during preadolescence, namely: working memory, assessed through the *n*-back task [14,31] and inattentiveness, assessed through the Attentional Network Test (ANT) [32,33]. Both tests have been validated as measures of neuropsychological development [32,34,35].

Working memory and attention are essential for learning and school attainment [12,13]. Working memory functions permit the maintenance and manipulation of information over short periods of time, and complex working memory requires continued, effective processing of information held in working memory stores. It underlies many other aspects of cognition, including learning, problem solving, reasoning, mathematics, and language comprehension [31]. Attention includes processes such as selectively attending to specific stimuli, focusing for prolonged periods on a task or incoming stimuli, or regulating and monitoring actions [14,36].

The experimental tasks were created for the project using the psychology experiment computer program E-Prime version 2.0 (Psychology Software Tools) and have been previously described [34,36,37].

In the *n*-back task, children observed a series of stimuli presented in the centre of the laptop's screen, and they were instructed to press a specific keyboard button whenever a given stimulus was the same as the one presented n trials previously (1-, 2-, and 3-back). Participants completed 3 blocks (1-, 2-, and 3-back) for each stimulus. Each block consisted of 25 trials, and blocks were separated by a short break (5 to 20 seconds). The first 3 trials of each block were never a target, and 33% of stimuli of the following trials were targets. The completion of a target was followed by a motivational sound ("woo hoo!") and a smiling face [37].

In the ANT, the screen showed a row of 5 yellow fish appearing either above or below a fixation point. Children were invited to "feed" the central fish as quickly as possible by pressing either the right or the left arrow key depending on the direction in which the target fish was pointing while ignoring the flanker fish, which pointed in either the same (congruent) or opposite (incongruent) direction than the middle fish. Visual signals informed about the approach of the target only (alerting cue) or about the approach of the target as well as its location (orienting cue) [38]. Each correct answer was followed by a simple animation sequence (the target fish blowing bubbles) and a recorded sound ("woo hoo!"). Incorrect responses were followed by a single tone and no animation of the fish [32]. A session consisted of 16 practice trials and 4 experimental blocks, each with 32 trials (128 trials in total) [36].

For the current study, we selected the following specific indicators because they showed little learning effect and an incremental growth in the repeated measurements during the study period. For the *n*-back test, we selected 2 loads (2-back and 3-back) and the numbers stimuli [26]. The 2-back test (or working memory hereafter) predicts general mental abilities, whereas the 3-back test (or complex working memory hereafter) is more complex to perform for children and could predict superior functions such as fluid intelligence [39]. All sets of *n*-back tests started with colours as a training phase and followed by the number stimuli. We evaluated the *n*-back parameter *d* prime (d'), a measure of detectability, obtained by subtracting the normalised false alarm rate from the hit rate: $(Z_{\text{hit rate}} - Z_{\text{false alarm rate}}) \times 100$. A higher d' indicates more accurate test performance, i.e., better working memory or complex working memory performance. For the ANT, we selected the hit reaction time standard error (HRT-SE). This is a measure of response speed consistency for correct responses throughout the test [40]. As such, a higher HRT-SE indicates a highly variable reaction, thus more inattentiveness (i.e., poorer sustained attention).

Covariates

We collected questionnaire information from parents about maternal and paternal education, marital status, occupation, family origin, gestational age at delivery, birth weight of the child, smoking during pregnancy, breastfeeding, siblings, adoption, and use of computer games. Area-level socioeconomic position (SES) at school and at home was derived using the

neighbourhood socioeconomic vulnerability index (area-level SES) at census tract level (median area of 0.08 km^2), a combined measure of 21 indicators covering 4 main dimensions: socioeconomic vulnerability, sociodemographic vulnerability, housing vulnerability, and subjective perception of vulnerability [41]. We measured height and weight and defined overweight and obesity using standard procedures [42]. We also collected information about the type of school (public/private) and assessed school educational quality as the sum of the school level (low, middle, high score) obtained in the basic competences of languages (Spanish and Catalan) and maths in the 2010/2011 Programme for International Student Assessment [43], which was self-reported by schools. Parents also completed the Strengths and Difficulties Questionnaire (SDQ) on child behavioural problems [44].

We accounted for annual average traffic-related air pollution (TRAP) levels outdoors and indoors at school by deriving a TRAP index consisting of 2 main traffic-related air pollutants in Barcelona, namely elemental carbon (EC) and NO_2 [45–47]. In parallel to noise measurements, these air pollutants were measured simultaneously indoors (in the classroom) and outdoors (in the courtyard) for each school pair. Measurements were carried out over two 1-week campaigns, which corresponded to the cold and warm seasons. EC was obtained with thermal-optical analyses [45] of the particulate matter <2.5 μm filters (high volume samples, quartz microfiber filters for sampling) deployed from 9 AM to 5 PM. Weekly NO_2 concentrations were measured with passive NO_2 samplers (NO_2 diffusion tube, Gradko International, United Kingdom). Annual concentrations were obtained by averaging the two 1-week measures after temporal adjustment by the ratio of the annual average to the weekly concentrations measured at a fixed background air quality monitoring station in Barcelona [47]. Individual exposure to traffic-related air pollution at home was estimated as annual average NO_2 levels at the geocoded postal addresses using land use regression models for the pollutants, as explained elsewhere [28].

Statistical analyses

We included a total of 2,680 (92.5%) children with complete data (i.e., with information at least for 1 outcome and school noise exposure and with data on age, sex, maternal education, and TRAP exposure), representing 9,984 (93.6%) tests. To account for the multilevel nature of the data, we used linear mixed effects models with the 4 repeated cognitive measures as outcomes (for each 2-back, 3-back, and HRT-SE), each noise exposure variable (one for each separate model) as a fixed effect predictor and child and school as nested random effects. To model the changes in the 12-month cognitive development associated with noise exposure, we further included an interaction term between age and the studied noise exposure variable [26]. As part of this model, we also reported the baseline effect (visit 1), namely the cross-sectional association between noise exposure and the cognitive outcome before the evaluation of the 12-month change in cognitive development. All models were further adjusted for potential confounders, as previously used [26]: age (years), sex (girl versus boy), maternal education (none/primary/secondary/university) as an indicator of household SES, urban vulnerability index as an indicator of residential neighbourhood-level SES, and outdoor or indoor TRAPs for models using outdoor or indoor noise at school, respectively, and for NO_2 at home for models using noise at home.

Moreover, we carried out stratified analyses to evaluate whether the trajectories of the cognitive outcomes over 1 year differed in children attending schools with low and high outdoor and indoor road traffic noise levels, using a cutoff of 55 dB for low/high outdoor noise levels, according to the EU definitions [1], and a cutoff of 30 dB for indoor noise according to the WHO guidelines for classroom noise levels [48]. The stratified models were adjusted for the same covariates of the linear mixed effects models described in the previous paragraph.

Sensitivity analyses

To assess the robustness of the evaluated associations to residual confounding, we further adjusted the main models for (one at a time) type of school (public versus private), paternal education (none/primary/secondary/university), foreign origin (child or the 2 parents born outside Spain/the child and at least 1 parent born in Spain), marital status (married or stable couple/others), overweight (non-overweight: body mass index < 85th percentile/overweight or obese: body mass index ≥ 85th percentile of WHO definition) [42], computer games during weekends (2 hours or less/>2 hours), siblings (Yes/No), adoption (Yes/No), smoking during pregnancy (Yes/No), preterm birth (<37 weeks/≥37 weeks), birth weight (<2.5 kg/≥2.5 kg), breastfeeding (Yes/No), socioeconomical vulnerability index at school (range: 0 to 1), school education quality (range: 0 to 6), and behavioural problems (range: 0 to 32). We also controlled for the paired design by including the school pair as a random effect (upon request during peer review). We also adjusted the studied associations at school for exposure to road traffic noise at home in the subsample living within the geographical limits covered by the 2012 Strategic Noise Map of Barcelona ($n = 2346$).

To compare our outdoor noise results with previous studies, which used modelled noise levels at street level instead of measurements, and to further assess the representativeness of our measurements for long-term exposure (i.e., annual averages), we repeated the main analyses by replacing the measured noise indicator (LAeq, street) by the modelled annual average level of road traffic noise for the daytime (Lday) estimated at the same location ($n = 34$ schools).

We reported the estimated change in the evaluated outcomes for a 5 dB, 50 events, and 10% increase in the noise level, number of noise events, and intermittency ratio, respectively. The statistical significance level used was $p < 0.05$, 2-sided. Analyses were performed with R statistical package (Version 3.4.2, R Foundation for Statistical Computing, Vienna, Austria) and Stata (Release 14. College Station, TX: StataCorp LP, United States).

Results

As shown in Table 1, a total of 2,508, 2,563, 2,493, and 2,420 children participated in visits 1 to 4, respectively, which represented a total participation of 2,680 children and 9,984 repeated outcome measurements. The median (percentile 25, percentile 75: p25; p75) age of the children in visit 1 was 8.5 (7.8; 9.3) years, and in visit 4, it was 9.4 (8.7; 10.2) years. During the 12-month period of the study, children's median (p25; p75) working memory (detectability) increased from 221 (131; 363) to 263 (153; 392) points, complex working memory

Table 1. Median (percentiles 25 and 75) of children's age and cognitive outcomes (working memory and inattentiveness) at each of the 4 repeated visits.

Visit	*n*	Age	2-back numbers (d')[a]	3-back numbers (d')[b]	ANT HRT-SE (ms)[c]
		p50 (p25, p75)	p50 (p25, p75)	p50 (p25, p75)	p50 (p25, p75)
1	2,508	8.5 (7.8, 9.3)	221 (131, 363)	112 (59, 171)	267 (201, 337)
2	2,563	8.7 (8.0, 9.5)	221 (131, 392)	123 (59, 190)	248 (184, 317)
3	2,493	9.1 (8.4, 9.8)	235 (131, 392)	129 (59, 190)	243 (181, 315)
4	2,420	9.4 (8.7, 10.2)	263 (153, 392)	129 (64, 212)	223 (162, 291)

[a]Working memory: 2-back number stimuli (d').

[b]Complex working memory: 3-back number stimuli (d'). d': detectability, a higher value indicates better working memory.

[c]Inattentiveness: HRT-SE (ms) of the Attention Network Task, a higher value indicates greater inattentiveness.

ANT, Attentional Network Test; HRT-SE, hit reaction time standard error.

https://doi.org/10.1371/journal.pmed.1004001.t001

(detectability) from 112 (59; 171) to 129 (64; 212) points, and inattentiveness (HRT-SE) decreased from 267 (201; 337) to 223 (162; 291) milliseconds.

The school- and individual-level characteristics of the study sample are presented in Table 2. Among others, a total of 49.0% of the sample were girls, 58.7% of the children had mothers, and 53% had fathers with high educational level (i.e., university) and the mean (standard deviation: SD) neighbourhood socioeconomic vulnerability index at residential level was 0.4 (0.2) points. Out of the 38 participating schools, 50% were public schools. Schools' mean (SD) road traffic noise levels were the highest in the street (LAeq,street), and decreased in the playground (LAeq,playgr) and even further indoors (LAeq,in) and for individual exposure in the child's classroom (LAeq,in-indiv), respectively: 63.6 (6.3) dB, 53.5 (5.4) dB, 38.6 (5.2) dB, and 37.5 (4.1) dB. Schools' mean (SD) noise intermittency ratio in the street (IR, street) was higher than in the playground (IR, playgr) and indoors (IR, in), respectively: 52.8 (16.2) %, 20.8 (11.3) %, and 25.0 (13.2) %. Similarly, the mean (SD) total number of noise events in the street (NE, street) was greater than in the playground (NE, playgr) and indoors (NE, in), respectively: 178.1 (59.6), 92.7 (49.0), and 102.6 (58.4) events. The same patterns were observed for school-outdoor versus school-indoor air pollution levels of EC, respectively: 1.5 (0.7) $\mu g/m^3$ and 1.4 (0.6) $\mu g/m^3$ and NO_2, respectively: 48.2 (13.2) $\mu g/m^3$ and 31.6 (13.1) $\mu g/m^3$. School's mean (SD) road traffic noise levels based on modelled estimates in the street was 65.6 (6.5) dB. Home-outdoor mean (SD) exposure to average road traffic noise (Lden) was 63.8 (7.8) dB and to NO_2 it was 54.6 (17.9) $\mu g/m^3$. Bivariate analyses (S2 Table) stratified by the median of LAeq,street at schools showed no statistically significant differences across percentage of public schools, neighbourhood socioeconomic vulnerability index at school or working and complex working memory at baseline. As main differences, schools with LAeq,street below the median (<63.5 dB) compared to those equal or above the median had less children with inattentiveness at visit 1 [median (interquartile range, IQR) = 262.3 (134.5) versus 271.9 (142.1), p-value < 0.012, Kruskal–Wallis test] and behavioural problems [median (IQR) = 7.0 (7.0) versus 8.0 (7.0), p-value < 0.001, Kruskal–Wallis test], had more girls (62.9% versus 54.1%, p-value < 0.001, χ^2 test), schools with higher educational quality [median (IQR) = 5.0 (3.0) versus 3.0 (2.0), p-value < 0.001, Kruskal–Wallis test], mothers with higher educational level (57.1% versus 48.7%, < 0.001, χ^2 test) and similarly for partners, and lower neighbourhood socioeconomic vulnerability index at home [median (IQR) = 0.4 (0.4) versus 0.5 (0.3), p-value < 0.001, Kruskal–Wallis test].

Pearson correlations between school-noise indicators were generally low or moderate (S3 Table), with some exceptions of high correlations between outdoor average noise levels (LAeq, street versus LAeq,playgr, r = 0.74, p-value < 0.001), between playground indicators of intermittency ratio and number of noise events (IR,playgr versus NE,playgr, r = 0.82, p-value < 0.001), and between indoor indicators of intermittency ratio and number of noise events (IR,in versus NE,in, r = 0.83, p-value < 0.001). Correlations between street and playground (i.e., outdoor) noise indicators were higher than between outdoor and indoor noise indicators [r range (p-value), respectively = 0.32 (0.059) to 0.74 (<0.001) versus 0.14 (0.426) to 0.44 (<0.001)], except for IR,street which exhibited, overall, low correlation with other indicators. Regarding air pollution, the correlation between outdoor and indoor TRAPs was high (r = 0.79, p-value < 0.001). The correlation between outdoor TRAP and outdoor noise indicators at school was moderately high with LAeq [r range (p-value): r = 0.68 (<0.001) to 0.70 (<0.001)], moderate with NE [r range (p-value): 0.40 (0.014) to 0.41 (0.012)] and low with IR [r range (p-value): −0.15 (0.384) to 0.28 (0.094)]. Similar magnitudes were observed for the correlation between indoor TRAP and indoor noise indicators in the classroom. The correlation between home-outdoor average noise levels (Lden) and NO_2 was r = 0.33 (p-value < 0.001). Finally, there was no correlation between outdoor exposure to noise at

Table 2. School- and individual-level characteristics of the study sample.

Variables	Mean (SD) or %
School-level variables (*n* = 38)	
Type of school, public	50.0%
School socioeconomic vulnerability index (*n*)	0.5 (0.2)
School education quality (PISA 2012, *n*)	3.9 (1.6)
Outdoor average noise level (LAeq, dB), street	63.6 (6.3)
Outdoor average noise level (LAeq, dB), playground	53.5 (5.4)
Indoor average noise level (LAeq, dB)	38.6 (5.2)
Outdoor noise intermittency ratio (IR, %), street	52.8 (16.2)
Outdoor noise intermittency ratio (IR, %), playground	20.8 (11.3)
Indoor noise intermittency ratio (IR, %)	25.0 (13.2)
Outdoor noise events (NE, *n*), street	178.1 (59.6)
Outdoor noise events (NE, *n*), playground	92.7 (49.0)
Indoor noise events (NE, *n*)	102.6 (58.4)
Outdoor elemental carbon level (EC, $\mu g/m^3$)	1.5 (0.7)
Indoor elemental carbon level (EC, $\mu g/m^3$)	1.4 (0.6)
Outdoor nitrogen dioxide level (NO_2, $\mu g/m^3$)	48.2 (13.2)
Indoor nitrogen dioxide level (NO_2, $\mu g/m^3$)	31.6 (13.1)
Outdoor average noise level (Lday, dB) (modelled)[a]	65.6 (6.5)
Individual-level variables (*n* = 2,680)	
Age (years)	8.5 (1.4)
Girls	49.9%
Maternal education, university	58.7%
Paternal education, university	53.3%
Foreign origin (non-Spanish)	14.9%
Marital status, married	85%
Overweight, Yes	27.6%
Computer games weekend, >1 hour	70.4%
Siblings, Yes	79.0%
Adopted child, Yes	3.9%
Smoking during pregnancy, Yes	10.2%
Birth $\geq$ 37 weeks	92.1%
Birth weight $\geq$ 2.5 kg	90.0%
Breastfeeding, Yes	82.0%
Behavioural problems (SDQ)	8.4 (5.2)
Home socioeconomic vulnerability index (*n*)	0.4 (0.2)
Individual indoor average noise level in classroom (LAeq, dB)	37.5 (4.1)
Home-outdoor average noise level (Lden, dB) (modelled)[b]	63.8 (7.8)
Home-outdoor average NO_2 level ($\mu g/m^3$) (modelled)	54.6 (17.9)

[a] n = 34.

[b] n = 2,346.

Data are mean (standard deviation) or percentage (%).

LAeq: A-weighted equivalent noise levels; Lday: LAeq for the daytime (7 AM to 9 PM); Lden: LAeq for the 24 hours with 5 dB and 10 dB penalties for the evening (9 PM to 11 PM) and nighttime (11 PM to 7 AM), respectively; PISA: Programme for International Student Assessment; SDQ: Strengths and Difficulties Questionnaire.

https://doi.org/10.1371/journal.pmed.1004001.t002

school and at home for any of the noise indicators [r range (*p*-value): −0.03 (0.165) to 0.02 (0.287)].

Association of exposure to road traffic noise with working memory outcomes and inattentiveness

The adjusted models for the association between exposure to road traffic noise at school and home and working memory, complex working memory and inattentiveness are shown in Table 3. Unadjusted models are shown in S4 Table.

Associations with average noise levels: School and home. At baseline, i.e., cross-sectionally, we observed some statistically significant associations (S4 Table) between average noise levels at school and the cognitive outcomes, which disappeared in adjusted models (Table 3), except for an association between street average noise levels and complex working memory, β = −3.92 points (95% CI: −7.74, −0.09, *p*-value = 0.045) per 5 dB and a general tendency in the expected direction in the estimated magnitudes for school average noise levels in the street, playground, and indoors in the classroom.

Regarding the 12-month change, both in unadjusted (S4 Table) and adjusted models (see Table 3), school-outdoor average noise levels (both at street level and at the playground) were consistently associated with a slower development of working memory, complex working memory, and with a slower improvement of inattentiveness over 12 months, which reached statistical significance except for the association between playground noise and inattentiveness. For example, a 5-dB increase in street average noise levels was related to a 12-month change of −4.83 points (95% CI: −7.21, −2.45, *p*-value < 0.001) in 2-back detectability, −4.01 points (95% CI: −5.91, −2.10, *p*-value < 0.001) in 3-back detectability, and 2.07 ms (95% CI: 0.37, 3.77, *p*-value = 0.017) in HRT-SE.

Table 3. Estimated effect (β) and 95% confidence intervals (95% CI) in working memory outcomes and inattentiveness at baseline and their 12-month change in association to school and home exposure to road traffic noise (*n* = 2,680 children, 9,984 repeats).

Road traffic noise indicators	Working memory (2-back numbers, *d'*)				Complex working memory (3-back numbers, *d'*)				Inattentiveness (Attention Network Task, HRT-SE[ms])			
	Baseline, β (95% CI)	*p*-value	12-month change, β (95% CI)	*p*-value	Baseline, β (95% CI)	*p*-value	12-month change, β (95% CI)	p-value	Baseline, β (95% CI)	*p*-value	12-month change, β (95% CI)	*p*-value
SCHOOL (MEASURED)												
Average level (LAeq, per 5 dB)												
Street	−1.98 (−6.27, 2.32)	0.367	−4.83 (−7.21, −2.45)	<0.001	−3.92 (−7.74, -0.09)	0.045	−4.01 (−5.91, −2.10)	<0.001	4.22 (−1.45, 9.90)	0.145	2.07 (0.37, 3.77)	0.017
Playground	−0.66 (−5.81, 4.49)	0.801	−3.68 (−6.76, -0.61)	0.019	−3.62 (−8.30, 1.06)	0.130	−4.41 (−6.89, −1.94)	<0.001	5.64 (−1.04, 12.32)	0.098	1.99 (−0.20, 4.17)	0.075
Indoor	−2.65 (−7.79, 2.48)	0.311	0.14 (−3.38, 3.66)	0.937	−1.94 (−6.52, 2.63)	0.405	−0.44 (−3.25, 2.38)	0.762	5.77 (−0.03, 11.58)	0.051	1.11 (−1.37, 3.59)	0.381
Individual indoor	0.02 (−4.17, 4.20)	0.994	−1.95 (−5.61, 1.71)	0.296	0.00 (−3.53, 3.52)	0.999	−0.80 (−3.78, 2.19)	0.601	1.28 (−1.52, 4.07)	0.371	2.41 (0.00, 4.81)	0.050
Intermittency ratio (per 10%)												
Street	1.03 (−1.53, 3.58)	0.431	0.71 (−1.29, 2.71)	0.486	1.34 (−1.08, 3.75)	0.278	−0.66 (−2.25, 0.94)	0.419	−2.08 (−5.38, 1.21)	0.216	0.43 (−1.00, 1.87)	0.554
Playground	−0.72 (−4.55, 3.10)	0.710	−2.59 (−5.45, 0.27)	0.075	−3.07 (−6.45, 0.31)	0.075	−3.42 (−5.71, −1.13)	0.003	−0.31 (−5.36, 4.74)	0.904	3.76 (1.71, 5.81)	<0.001
Indoor	1.65 (−1.67, 4.97)	0.330	−2.27 (−4.67, 0.13)	0.063	0.15 (−2.89, 3.19)	0.923	−2.76 (−4.66, -0.85)	0.005	−1.45 (−5.46, 2.56)	0.479	3.05 (1.34, 4.76)	<0.001
Number of events (per 50)												
Street	0.58 (−3.21, 4.36)	0.765	−4.38 (−7.08, −1.67)	0.002	−1.08 (−4.61, 2.44)	0.547	−3.99 (−6.16, -1.82)	<0.001	−0.25 (−5.15, 4.65)	0.920	2.13 (0.18, 4.08)	0.032
Playground	−0.68 (−4.84, 3.48)	0.749	−2.68 (−5.71, 0.35)	0.082	−3.04 (−6.85, 0.78)	0.119	−3.03 (−5.46, -0.61)	0.014	−0.77 (−6.70, 5.16)	0.800	2.22 (0.07, 4.37)	0.043
Indoor	2.04 (−2.39, 6.46)	0.367	−2.72 (−5.45, 0.01)	0.051	−1.18 (−5.14, 2.78)	0.559	−3.22 (−5.39, −1.04)	0.004	−1.27 (−6.55, 4.01)	0.636	3.15 (1.20, 5.11)	0.002
SCHOOL & HOME (MODELLED)												
Average level (LAeq, per 5 dB)												
School street, Lday[a]	−2.98 (−8.78, 2.82)	0.314	−6.17 (−8.84, −3.49)	<0.001	−4.15 (−9.14, 0.85)	0.104	−4.91 (−7.03, −2.78)	<0.001	3.42 (−4.45, 11.29)	0.394	2.81 (0.89, 4.74)	0.004
Home street, Lden[b]	0.00 (−2.39, 2.40)	0.997	1.35 (−0.83, 3.53)	0.224	0.75 (−1.11, 2.61)	0.430	0.40 (−1.32, 2.12)	0.648	0.72 (−1.25, 2.68)	0.474	−0.52 (−2.06, 1.01)	0.505

[a] $n = 34$.

[b] $n = 2{,}346$.

Linear mixed models adjusted for age, sex, maternal education, socioeconomical vulnerability index at home, outdoor or indoor TRAP at school or outdoor NO_2 at home, respectively, for models with the corresponding noise indicators (i.e., outdoors or indoors at school or outdoors at home). Child and school included as nested random effects. The 12-month change models include the term age × corresponding noise indicator to estimate the change; *d'*: detectability, a higher value indicates better working memory; HRT-SE: hit reaction time standard error, a higher value indicates greater inattentiveness; LAeq: A-weighted equivalent noise levels; Lday: LAeq for the daytime (7 AM to 9 PM); Lden: LAeq for the 24 hours with 5 dB and 10 dB penalties for the evening (9 PM to 11 PM) and nighttime (11 PM to 7 AM), respectively; NO_2: nitrogen dioxide; TRAP: traffic-related air pollution.

https://doi.org/10.1371/journal.pmed.1004001.t003

Regarding school-indoor average noise levels, individual average noise exposure in classrooms was associated with a slower improvement of inattentiveness in unadjusted models and was borderline significant in adjusted models and of similar magnitude to results with outdoor levels: β = 2.41 ms (95% CI: 0.00, 4.81, *p*-value = 0.050) per 5 dB (adjusted model). In contrast, no association with indoor noise levels was present for working memory and complex working memory (e.g., −0.80 points; 95% CI: −3.78, 2.19, *p*-value = 0.601, in the 12-month change in 3-back detectability per 5 dB). Indoor average noise levels based on measurements in 1 classroom were not associated with any of the outcomes, although the estimated magnitude was in the expected direction (e.g., 1.11 ms (95% CI: −1.37, 3.59, *p*-value = 0.381) in the 12-month change in HRT-SE per 5 dB) (Table 3).

There was no association between home-outdoor average noise levels (Lden) and any of the cognitive outcomes.

Associations with noise fluctuation indicators at school. Both school-outdoor and indoor exposure to noise intermittency and to noise events were associated with slower development in the cognitive outcomes in children over 12 months in unadjusted (S4 Table) and adjusted models (Table 3). Some associations were observed at baseline in unadjusted models (S4 Table); however, they disappeared after further adjustment (Table 3).

Regarding intermittency ratio (Table 3), a 10% increment in playground IR was associated with a 12-month change of −2.59 points (95% CI: −5.45, 0.27, *p*-value = 0.075) in 2-back detectability, −3.42 points (95% CI: −5.71, −1.13, *p*-value = 0.003) in 3-back detectability, and 3.76 ms (95% CI: 1.71, 5.81, *p*-value < 0.001) in HRT-SE, and a 10% increment in indoor IR was associated with a 12-month change of −2.27 points (95% CI: −4.67, 0.13, *p*-value = 0.063) in 2-back detectability, −2.76 points (95% CI: −4.66, −0.85, *p*-value = 0.005) in 3-back detectability, and 3.05 ms (95% CI: 1.34, 4.76, *p*-value < 0.001) in HRT-SE. Street IR was not associated with any of the outcomes.

Associations were also observed for street, playground, and indoor NE with a slower development in 2-back and 3-back detectability and a slower improvement of inattentiveness over 12 months, which were statistically significant except for playground and indoor NE with 2-back detectability (Table 3). For instance, increments in exposure of 50 noise events at street, playground, and indoor level were associated, respectively, with a 12-month change of −4.38 points (95% CI: −7.08, −1.67, *p*-value = 0.002), −2.68 (95% CI: −5.71, 0.35, *p*-value = 0.082), and −2.72 (95% CI: −5.45, 0.01, *p*-value = 0.051) in 2-back detectability, and −3.99 points (95% CI: −6.16, −1.82, *p*-value < 0.001), −3.03 (95% CI: −5.46, −0.61, *p*-value = 0.014), and −3.22 (95% CI: −5.39, −1.04, *p*-value = 0.004) in 3-back detectability, and 2.13 ms (95% CI: 0.18, 4.08, *p*-value = 0.032), 2.22 (95% CI: 0.07, 4.37, *p*-value = 0.043), and 3.15 (95% CI: 1.20, 5.11, *p*-value = 0.002) in HRT-SE.

Trajectories of working memory and inattentiveness at schools above and below recommended noise levels

As shown in Fig 1, children who attended schools with high road traffic noise at street level (≥55 dB) had a slower development of working memory, complex working memory, and a slower improvement of inattentiveness over 12 months than those attending quieter schools in adjusted models. Similar trends with slightly weaker differences between groups were observed for schools exposed to high noise at the playground. Finally, children who attended schools with high road traffic noise in the classroom (≥30 dB) had a slower improvement of inattentiveness over 12 months than those who attended schools with quieter classrooms.

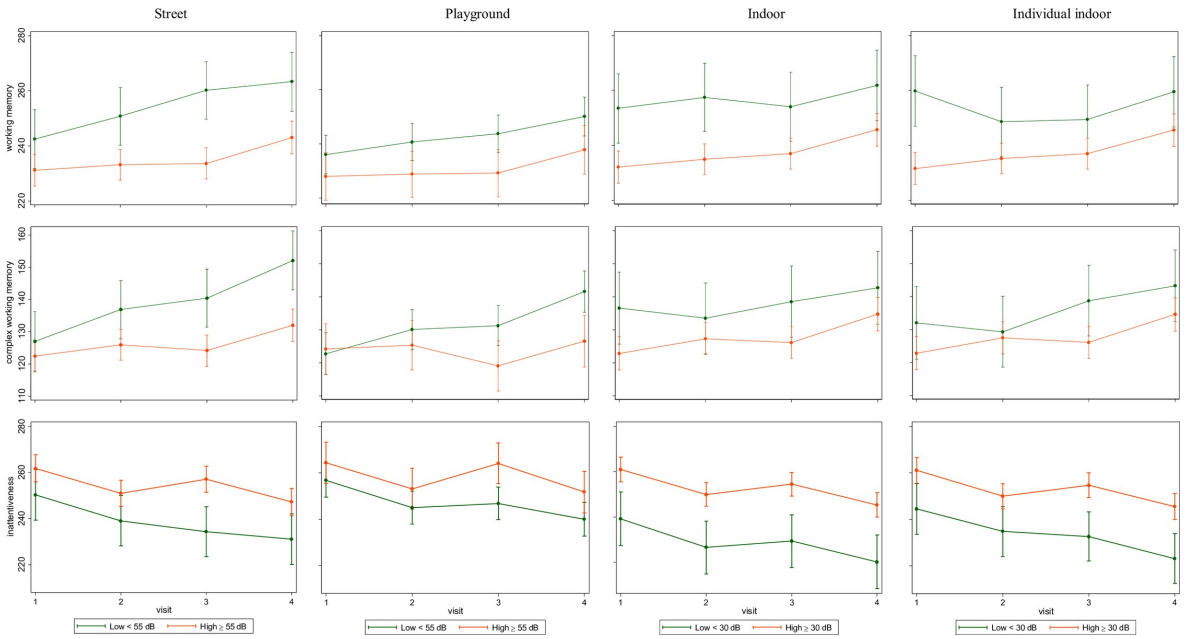

Fig 1. Annual trajectories of working memory, complex working memory, and inattentiveness in children attending schools with low and high average road traffic noise levels (LAeq, dB) outdoors in street and playground or indoors in 1 classroom (indoor level) or in each child's classroom considering change of room between years (individual indoor level). Y axis: point estimate (beta coefficient), error bars (95% confidence intervals). Predictions for working memory (2-back number stimuli, *d'*), complex working memory (3-back number stimuli, *d'*), and inattentiveness (HRT-SE, ms) adjusted at the means of age, sex, corresponding road traffic noise indicator, age* road traffic noise indicator, maternal education, socioeconomical vulnerability index at home and outdoor or indoor TRAP at school for models including outdoor or indoor noise levels, respectively. Child and school included as nested random effects. *d'*: detectability, a higher value indicates better working memory; HRT-SE: hit reaction time standard error, a higher value indicates greater inattentiveness; LAeq: A-weighted equivalent noise levels; TRAP: traffic-related air pollution.

https://doi.org/10.1371/journal.pmed.1004001.g001

Sensitivity analyses

When using modelled (Lday) instead of measured average road traffic noise levels at the school street (see bottom of Table 3), we observed the same consistent associations with a slower development of all the cognitive outcomes over 12 months.

Associations between the different road traffic noise indicators at school and the development of cognitive outcomes were robust to additional adjustment for other potential confounders (S5–S7 Tables) and to the inclusion or exclusion of traffic-related air pollution in the adjustment sets. Although in certain adjustment sets the associations of indoor intermittency ratio and indoor number of noise events with working memory gained significance and the association between individual indoor LAeq and inattentiveness lost significance, the estimated magnitudes of effect remained unchanged.

Discussion

In the current study, exposure to road traffic noise at school, but not at home, was associated with a slower development of working memory and of complex working memory, and with a slower improvement of inattentiveness over 1 year in schoolchildren. Exposure to road traffic noise (at school and home) was not associated with the cognitive outcomes cross-sectionally (i.e., at the baseline visit). Both the school-outdoor average level and fluctuation characteristics

of noise were associated with a deceleration in the studied cognitive outcomes. In contrast, in the classroom, the noise fluctuation characteristics were more robustly associated with all cognitive outcomes, whereas average levels were only associated with greater inattentiveness. Finally, children attending schools exposed to outdoor road traffic noise levels $\geq$55 dB had a slower development of working and complex working memory and greater inattentiveness and those with classroom levels $\geq$30 dB had greater inattentiveness over 12 months, compared to children exposed to lower outdoor ($<$55 dB) and indoor ($<$30 dB) noise levels, respectively.

Associations with working memory (2-back detectability) were generally in the expected direction but only reached statistical significance in relation to street-level indicators (average noise and number of events). This could relate to the fact that the 2-back detectability test was less challenging for children, thus potentially less sensitive to capture changes in development than complex working memory (3-back detectability) for which associations with noise indicators were statistically significant.

As we observed, while individual exposure to the annual average noise level in the classroom was only associated with inattentiveness, exposure to intermittent noise and to a greater number of noise events in the classroom was associated both with greater inattentiveness and slower complex working memory and also marginally with slower working memory development. These findings support the hypothesis that the noise characteristics beyond the average noise level, i.e., its fluctuation, might be more relevant for children's neurodevelopment in the classroom. They also support the importance of carrying out detailed indoor noise exposure assessment in studies of the cognitive effects of noise, to move closer to the personal exposure inside the classroom. In other words, the peaks of road traffic noise that propagate into the classroom (and their frequency) could be further disruptive for children's working memory and attention development during concentration at school even when the average noise level in the classroom is lower and may only affect attention. The relevance of the same fluctuation characteristics of noise was previously observed for the effects of transportation noise on cardiovascular outcomes [24]. These results suggest that noise fluctuation should be further investigated and that policy recommendations to protect children's health may have to consider noise fluctuation in addition to average noise levels. We did not observe associations between exposure to road traffic noise at home and the studied cognitive outcomes, which is in line with the only previous study assessing both school and home exposure [20,21]. This could suggest that exposure to noise at school, rather than at home, may be more detrimental by affecting vulnerable windows of concentration and learning processes [8]. Another complementary explanation may relate to a greater degree of exposure misclassification for noise exposure at home, as we could only assign outdoor average noise levels at the home address, as commonly done in previous studies. Moreover, we could not estimate fluctuation measures, which were relevant indicators associated with cognitive development at school.

We observed very few associations between exposure to road traffic noise at school and the cognitive outcomes at baseline. The baseline results represent the cross-sectional association between exposure to noise and cognitive outcomes during the first visit. The general lack of associations at baseline could partly indicate that the children's long-term noise exposure years before baseline was not captured with our exposure assessment, given that we did not have historical information about exposure of children in previous schools, although a 98% of children had attended the same school for at least 1 year. Alternatively, the more consistent associations with the 12-month development might respond to the vulnerable window of effects of the studied cognitive functions, which develop significantly during primary school age [14–16]. Finally, inconsistencies in the baseline results could be partly inherent to the cross-sectional nature of the observation, and would be in line with the mixed findings of previous cross-sectional studies [17–23].

Comparison with previous literature

There is limited evidence for the association between road traffic noise and cognitive development in children, including working memory and attention, which are crucial for learning and school attainment [12,13] and which are actively developing during primary school age [14]. The few studies so far were all cross-sectional and used diverse outcome tests, which precludes direct comparison with the current study. The multicentric and cross-sectional RANCH study, carried out in 2,844 schoolchildren of 9 to 10 years of age in the Netherlands, Spain, and the UK in 2002, observed associations between annual average noise levels of road traffic noise outdoors at schools (daytime LAeq) and cognitive performance, in terms of reading comprehension and episodic memory [19,23]. In line with our study, they observed associations between school-outdoor road traffic noise levels at school, but not at home, and inattention in the 553 children in the Netherlands, based on the Switching Attention Test [20,21]. However, the RANCH project did not find associations with working memory, assessed with a modified version of The Search and Memory Task [19–23], or with sustained attention [19,20,22], assessed with a modified version of the Toulouse–Pieron test [19]. Finally, 2 small studies, one studying exposure to road traffic noise at home in 73 children and another one at 2 schools, found respectively no [17] and suggestive [18] associations with tests of attention. None of the aforementioned studies evaluated exposure to road traffic noise indoors in the classroom or noise fluctuation measures.

Biological mechanisms

Our results support the general hypothesis that childhood may be a vulnerable period in which external stimuli such as noise could affect the rapid cognitive development that is occurring [7]. There are several suggested pathways on how noise may affect cognitive development in children. One suggested pathway is that noise may affect cognition by impairing attention [8], which is in line with our findings with inattentiveness. It is also suggested that noise could directly impair children's cognitive abilities, or act indirectly through frustration, learned helplessness, increased arousal, or tuning out, which could impact performance and learning in the long term under repeated noise exposure [8]. The effects of noise on cognition are further supported by animal experiments showing that exposure to environmental noise levels (72 dB, aircraft, up to 4 days) increases oxidative stress in the frontal cortex [11] and subchronic exposure to white noise (4 hours/day, 100 dB, up to 30 days) leads to brain changes in occipital and prefrontal regions [9,10] and in the hippocampus [10], which are relevant for executive functions, learning, and memory. Overall, the suggested pathways and our results support the hypothesis that the effects of noise on cognition, here executive functions, such as working memory, and attention, may be greater during activities that involve concentration or attention and that may affect learning.

Strengths and limitations

The major strength of this study was the longitudinal design with repeated cognitive evaluations, which allowed to study for the first time the association between exposure to road traffic noise and the development in the cognitive functions of working memory and attention in schoolchildren. We also used validated computerised tests that could capture different dimensions of cognition and executive functions that are developing in primary schoolchildren [34]. Moreover, to our knowledge, this study performed the most extensive and detailed noise exposure assessment of road traffic noise to date, including exposure in different microenvironments at school (at street level, playground, and indoors in the classroom), assessing not only average noise levels but also fluctuation characteristics of noise at school (i.e., intermittency

ratio and number of noise events), and exposure to average road traffic noise levels outdoors at home. Furthermore, our road traffic noise measurement protocol yielded representative estimates of the annual average noise levels during the study period, as shown by the high correlation between the annual average noise levels from Barcelona's Strategic noise map published in 2012 and our school measurements at the street level in 2012 (Spearman's rank r = 0.86, Pearson's r = 0.85), the consistent results found with these 2 exposures (modelled and measured) and also the high intraclass correlation (ICC = 0.94) between indoor noise measurements in the 2 campaigns, which were performed 6 months apart.

Among the limitations, our study did not assess past noise exposure. However, all children had attended the same school at least for 6 months and 98% had attended for more than 1 year. In turn, it is unlikely that road traffic noise levels were different during the previous years, as it would require drastic changes in traffic. The lack of association between exposure to noise at home and the cognitive outcomes could partly relate to the greater exposure misclassification expected with the residential assignment of modelled noise levels at the home address. Another limitation is the potential lack of external validity of our results, given that 58% of the participating mothers had a university degree compared to a 50% of women between 25 and 39 years of age in the region [49]. However, maternal education did not seem to determine participation, given that the participation rate was independent of the school-area socioeconomic level (Spearman's rank correlation = −0.09, p-value = 0.61). Besides, we could not study children with specific needs, as this would require a larger sample size to have statistical power. Furthermore, residual confounding is always a possibility, although we adjusted for multiple potential confounders and results were robust to additional adjustments in sensitivity analysis. Finally, air pollution could be an important confounder of the studied association, given that it has been associated with impaired cognitive development [26]. In the current study, we adjusted our analyses for TRAP. Moreover, the correlation between TRAPs and the measured noise fluctuation indicators was moderate or low, which further supports that the observed associations between road traffic noise and deceleration in working memory and attention development were independent of air pollution.

Public health relevance

At least 1 in every 5 Europeans are exposed to high road traffic noise levels (Lden $\geq$ 55 dB) [1]. In Barcelona, more than half of the population are affected by road traffic noise levels above those recommended by the WHO (Lden $\geq$ 53 dB) [4,27], and in the current study, more than half of the schools were also exceeding the recommended WHO thresholds both outside and inside classrooms. The current findings suggest that exposure to road traffic noise is associated with a slower development of the cognitive functions of working memory and attention in primary schoolchildren of Barcelona, which are essential for learning and school attainment [12,13]. Given the expected large number of children exposed to road traffic noise at schools, particularly in urban areas, the application of policies to reduce road traffic noise at schools (outside and inside classrooms) could substantially benefit cognitive development, at least working memory and attention, and future health. Furthermore, the current findings add to the previous evidence about the adverse effects of school exposure to aircraft noise on other dimensions of children's cognition [5,8], and also to our previous findings for the associations of air pollution [26,50,51] with working memory and attention in the same cohort of children. Together, this evidence indicates that efficient interventions to protect the school environment should target transportation and consider not only cleaner air but also quieter school environments.

In conclusion, exposure to road traffic noise at school, but not at home, was associated with slower working memory, complex working memory, and attention development, in primary

school children. Associations were observed both for school-outdoor average noise levels and noise fluctuation indicators, although in classrooms, noise fluctuation was more consistently associated with all cognitive outcomes than average noise levels. Finally, slower development of working memory, complex working memory, and attention was observed in children attending schools exposed to outdoor road traffic noise levels above ≥55 dB and to classroom levels ≥30 dB, compared to children exposed to lower outdoor (<55 dB) and indoor (<30 dB) noise levels, respectively. Further longitudinal studies are needed to replicate these findings in different populations and settings, to assess different microenvironments and noise fluctuation metrics and to study other cognitive functions developing over the first years of life.

Supporting information

S1 STROBE Checklist. STROBE Statement—Checklist of items that should be included in reports of cohort.
(PDF)

S1 Table. Percent or median (interquartile range) of paired school characteristics by low or high school nitrogen dioxide levels.
(PDF)

S2 Table. Main school- and individual-level characteristics of the study sample by the median of street-outdoor average noise levels at school (LAeq, dB).
(PDF)

S3 Table. Pearson correlations between noise and traffic-related air pollution levels at schools (*n* = 38) and at individual level (*n* = 2,680 children).
(PDF)

S4 Table. Estimated unadjusted effect (β) and 95% confidence intervals (95% CI) in cognitive outcomes at baseline and their 12-month change in association to school and home exposure to road traffic noise (*n* = 2,680 children, 9,984 repeats).
(PDF)

S5 Table. Unadjusted and additionally adjusted models for the association between school exposure to average noise levels (LAeq, per 5 dB change) and 12-month change of working memory, complex working memory, and inattentiveness. (*n* = 2,680 children, 9,984 repeats).
(PDF)

S6 Table. Unadjusted and additionally adjusted models for the association between school exposure to noise intermittency ratio (IR, per 10% change) and 12-month change of working memory, complex working memory, and inattentiveness (*n* = 2,680 children, 9,984 repeats).
(PDF)

S7 Table. Unadjusted and additionally adjusted models for the association between school exposure to number of noise events (NE, per 50 events) and 12-month change of working memory, complex working memory, and inattentiveness (*n* = 2,680 children, 9,984 repeats).
(PDF)

S1 Text. Estimation of the individual indoor noise levels in the classroom.
(PDF)

Acknowledgments

We acknowledge all the families participating in the study for their altruism and particularly to the schools Antoni Brusi, Baloo, Betània–Patmos, Centre d'Estudis Montseny, Col·legi Shalom, Costa i Llobera, Escola El Sagrer, Els Llorers, Escola Pia de Sarrià, Escola Pia Balmes, Escola Concertada Ramon Llull, Escola Lourdes, Escola Tècnica Professional del Clot, Ferran i Clua, Francesc Macià, Frederic Mistral, Infant Jesús, Joan Maragall, Jovellanos, La Llacuna del Poblenou, Lloret, Menéndez Pidal, Nuestra Señora del Rosario, Miralletes, Ramon Llull, Rius i Taulet, Pau Vila, Pere Vila, Pi d'en Xandri, Projecte, Prosperitat, Sant Ramon Nonat–Sagrat Cor, Santa Anna, Sant Gregori, Sagrat Cor Diputació, Tres Pins, Tomàs Moro, Torrent d'en Melis, and Virolai. We also acknowledge the ESCAPE project for the design and supervision modelling of air pollution. Furthermore, Xavier Mayoral did the technical development of the *n*-back test; Cecilia Persavento, Judit Gonzalez, Laura Bouso, and Pere Figueras contributed to the field work. Joan Forns contributed to the design and project management.

Author Contributions

Conceptualization: Maria Foraster, Jordi Sunyer.

Data curation: Maria Foraster, Mikel Esnaola, Mónica López-Vicente, Ioar Rivas, Mar Álvarez-Pedrerol.

Formal analysis: Maria Foraster, Mikel Esnaola.

Funding acquisition: Maria Foraster, Jordi Sunyer.

Investigation: Maria Foraster, Mónica López-Vicente, Ioar Rivas, Cecilia Persavento, Jordi Sunyer.

Methodology: Maria Foraster, Mikel Esnaola, Mónica López-Vicente, Ioar Rivas, Nuria Sebastian-Galles, Jesus Pujol, Payam Dadvand, Jordi Sunyer.

Project administration: Mar Álvarez-Pedrerol, Cecilia Persavento, Jordi Sunyer.

Resources: Mar Álvarez-Pedrerol.

Software: Mikel Esnaola.

Supervision: Jordi Sunyer.

Writing – original draft: Maria Foraster.

Writing – review & editing: Maria Foraster, Mikel Esnaola, Mónica López-Vicente, Ioar Rivas, Mar Álvarez-Pedrerol, Nuria Sebastian-Galles, Jesus Pujol, Payam Dadvand, Jordi Sunyer.

References

1. European Environment Agency. Environmental noise in Europe, 2020. EEA Report No 22/2019. 2020. Available from: https://op.europa.eu/publication/manifestation_identifier/PUB_THAL20003ENN.
2. Hänninen O, Knol AB, Jantunen M, Lim T-A, Conrad A, Rappolder M, et al. Environmental burden of disease in Europe: assessing nine risk factors in six countries. Environ Health Perspect. 2014; 122:439–46. https://doi.org/10.1289/ehp.1206154 PMID: 24584099
3. World Health Organization, Comission European, editors. Burden of disease from environmental noise: quantification of healthy life years lost in Europe. Copenhagen: World Health Organization, Regional Office for Europe; 2011.
4. WHO Regional Office for Europe. Environmental Noise Guidelines for the European Region (2018). Copenhagen: World Health Organization; 2018.

5. Clark C, Crumpler C, Notley AH. Evidence for Environmental Noise Effects on Health for the United Kingdom Policy Context: A Systematic Review of the Effects of Environmental Noise on Mental Health, Wellbeing, Quality of Life, Cancer, Dementia, Birth, Reproductive Outcomes, and Cognition. Int J Environ Res Public Health. 2020;17. https://doi.org/10.3390/ijerph17020393 PMID: 31936110

6. Basner M, Babisch W, Davis A, Brink M, Clark C, Janssen S, et al. Auditory and non-auditory effects of noise on health. Lancet. 2014; 383:1325–32. https://doi.org/10.1016/S0140-6736(13)61613-X PMID: 24183105

7. Stansfeld S, Clark C. Health Effects of Noise Exposure in Children. Curr Environ Health Rep. 2015; 2:171–8. https://doi.org/10.1007/s40572-015-0044-1 PMID: 26231366

8. Clark C, Paunovic K. WHO Environmental Noise Guidelines for the European Region: A Systematic Review on Environmental Noise and Cognition. Int J Environ Res Public Health. 2018; 15:285. https://doi.org/10.3390/ijerph15020285 PMID: 29414890

9. Loganathan S, Rathinasamy S. Alteration in Memory and Electroencephalogram Waves with Subacute Noise Stress in Albino Rats and Safeguarded by Scoparia dulcis. Pharmacogn Mag. 2016; 12: S7–S13. https://doi.org/10.4103/0973-1296.176119 PMID: 27041862

10. Manikandan S, Padma MK, Srikumar R, Jeya Parthasarathy N, Muthuvel A, Sheela DR. Effects of chronic noise stress on spatial memory of rats in relation to neuronal dendritic alteration and free radical-imbalance in hippocampus and medial prefrontal cortex. Neurosci Lett. 2006; 399:17–22. https://doi.org/10.1016/j.neulet.2006.01.037 PMID: 16481110

11. Kröller-Schön S, Daiber A, Steven S, Oelze M, Frenis K, Kalinovic S, et al. Crucial role for Nox2 and sleep deprivation in aircraft noise-induced vascular and cerebral oxidative stress, inflammation, and gene regulation. Eur Heart J. 2018; 39:3528–39. https://doi.org/10.1093/eurheartj/ehy333 PMID: 29905797

12. Alloway TP, Alloway RG. Investigating the predictive roles of working memory and IQ in academic attainment. J Exp Child Psychol. 2010; 106:20–9. https://doi.org/10.1016/j.jecp.2009.11.003 PMID: 20018296

13. Spira EG, Fischel JE. The impact of preschool inattention, hyperactivity, and impulsivity on social and academic development: a review. J Child Psychol Psychiatry. 2005; 46:755–73. https://doi.org/10.1111/j.1469-7610.2005.01466.x PMID: 15972069

14. Anderson P. Assessment and development of executive function (EF) during childhood. Child Neuropsychol. 2002; 8:71–82. https://doi.org/10.1076/chin.8.2.71.8724 PMID: 12638061

15. Østby Y, Tamnes CK, Fjell AM, Walhovd KB. Morphometry and connectivity of the fronto-parietal verbal working memory network in development. Neuropsychologia. 2011; 49:3854–62. https://doi.org/10.1016/j.neuropsychologia.2011.10.001 PMID: 22001853

16. Ullman H, Almeida R, Klingberg T. Structural Maturation and Brain Activity Predict Future Working Memory Capacity during Childhood Development. J Neurosci. 2014; 34:1592–8. https://doi.org/10.1523/JNEUROSCI.0842-13.2014 PMID: 24478343

17. Cohen S, Glass DC, Singer JE. Apartment noise, auditory discrimination, and reading ability in children. J Exp Soc Psychol. 1973; 9:407–22. https://doi.org/10.1016/S0022-1031(73)80005-8

18. Sanz SA, García AM, García A. Road traffic noise around schools: a risk for pupil's performance? Int Arch Occup Environ Health. 1993; 65:205–7. https://doi.org/10.1007/BF00381157 PMID: 8282419

19. Stansfeld SA, Berglund B, Clark C, Lopez-Barrio I, Fischer P, Ohrström E, et al. Aircraft and road traffic noise and children's cognition and health: a cross-national study. Lancet. 2005; 365:1942–9. https://doi.org/10.1016/S0140-6736(05)66660-3 PMID: 15936421

20. van Kempen E, van Kamp I, Lebret E, Lammers J, Emmen H, Stansfeld S. Neurobehavioral effects of transportation noise in primary schoolchildren: a cross-sectional study. Environ Health. 2010; 9:25. https://doi.org/10.1186/1476-069X-9-25 PMID: 20515466

21. van Kempen E, Fischer P, Janssen N, Houthuijs D, van Kamp I, Stansfeld S, et al. Neurobehavioral effects of exposure to traffic-related air pollution and transportation noise in primary schoolchildren. Environ Res. 2012; 115:18–25. https://doi.org/10.1016/j.envres.2012.03.002 PMID: 22483436

22. Clark C, Crombie R, Head J, van Kamp I, van Kempen E, Stansfeld SA. Does traffic-related air pollution explain associations of aircraft and road traffic noise exposure on children's health and cognition? A secondary analysis of the United Kingdom sample from the RANCH project. Am J Epidemiol. 2012; 176:327–37. https://doi.org/10.1093/aje/kws012 PMID: 22842719

23. Matheson M, Clark C, Martin R, van Kempen E, Haines M, Barrio IL, et al. The effects of road traffic and aircraft noise exposure on children's episodic memory: The RANCH Project. Noise Health 2010; 12: 244. https://doi.org/10.4103/1463-1741.70503 PMID: 20871179

24. Foraster M, Eze IC, Schaffner E, Vienneau D, Héritier H, Endes S, et al. Exposure to Road, Railway, and Aircraft Noise and Arterial Stiffness in the SAPALDIA Study: Annual Average Noise Levels and

Temporal Noise Characteristics. Environ Health Perspect. 2017; 125:097004. https://doi.org/10.1289/EHP1136 PMID: 28934719

25. Brink M, Schaeffer B, Vienneau D, Foraster M, Pieren R, Eze IC, et al. A survey on exposure-response relationships for road, rail, and aircraft noise annoyance: Differences between continuous and intermittent noise. Environ Int. 2019; 125:277–90. https://doi.org/10.1016/j.envint.2019.01.043 PMID: 30731377

26. Sunyer J, Esnaola M, Alvarez-Pedrerol M, Forns J, Rivas I, López-Vicente M, et al. Association between traffic-related air pollution in schools and cognitive development in primary school children: a prospective cohort study. PLoS Med. 2015; 12:e1001792. https://doi.org/10.1371/journal.pmed.1001792 PMID: 25734425

27. Font L, Gómez A, Oliveras L, Realp E, Borrell C. Soroll ambiental i salut a la ciutat de Barcelona (Environmental Noise and Health in the city of Barcelona). Barcelona, Spain: Public Health Agency of Barcelona; 2020. Available from: https://www.aspb.cat/documents/sorollambientalisalut-bcn2021/.

28. Wang M, Beelen R, Basagana X, Becker T, Cesaroni G, de Hoogh K, et al. Evaluation of Land Use Regression Models for NO2 and Particulate Matter in 20 European Study Areas: The ESCAPE Project. Environ Sci Technol. 2013; 47:4357–64. https://doi.org/10.1021/es305129t PMID: 23534892

29. European Commission Directive 2002/49/EC of the European Parliament and of the council of 25 June 2002 relating to the assessment and management of environmental noise. European Commission; 2002. p. 0012–0026.

30. Wunderli JM, Pieren R, Habermacher M, Vienneau D, Cajochen C, Probst-Hensch N, et al. Intermittency ratio: A metric reflecting short-term temporal variations of transportation noise exposure. J Expo Sci Environ Epidemiol. 2016; 26:575–85. https://doi.org/10.1038/jes.2015.56 PMID: 26350982

31. Vuontela V, Steenari M-R, Carlson S, Koivisto J, Fjällberg M, Aronen ET. Audiospatial and Visuospatial Working Memory in 6–13 Year Old School Children. Learn Mem. 2003; 10:74–81. https://doi.org/10.1101/lm.53503 PMID: 12551966

32. Rueda MR, Fan J, McCandliss BD, Halparin JD, Gruber DB, Lercari LP, et al. Development of attentional networks in childhood. Neuropsychologia. 2004; 42:1029–40. https://doi.org/10.1016/j.neuropsychologia.2003.12.012 PMID: 15093142

33. Rueda MR, Rothbart MK, McCandliss BD, Saccomanno L, Posner MI. Training, maturation, and genetic influences on the development of executive attention. Proc Natl Acad Sci U S A. 2005; 102:14931–6. https://doi.org/10.1073/pnas.0506897102 PMID: 16192352

34. Forns J, Esnaola M, López-Vicente M, Suades-González E, Alvarez-Pedrerol M, Julvez J, et al. The n-back test and the attentional network task as measures of child neuropsychological development in epidemiological studies. Neuropsychology. 2014; 28:519–29. https://doi.org/10.1037/neu0000085 PMID: 24819069

35. Thomason ME, Race E, Burrows B, Whitfield-Gabrieli S, Glover GH, Gabrieli JDE. Development of spatial and verbal working memory capacity in the human brain. J Cogn Neurosci. 2009; 21:316–32. https://doi.org/10.1162/jocn.2008.21028 PMID: 18510448

36. Suades-González E, Forns J, García-Esteban R, López-Vicente M, Esnaola M, Álvarez-Pedrerol M, et al. A Longitudinal Study on Attention Development in Primary School Children with and without Teacher-Reported Symptoms of ADHD. Front Psychol. 2017;8. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2017.00655.

37. López-Vicente M, Forns J, Suades-González E, Esnaola M, García-Esteban R, Álvarez-Pedrerol M, et al. Developmental Trajectories in Primary Schoolchildren Using n-Back Task. Front Psychol. 2016;7. Available from: https://www.frontiersin.org/article/10.3389/fpsyg.2016.00716.

38. Rueda MR, Checa P, Cómbita LM. Enhanced efficiency of the executive attention network after training in preschool children: Immediate changes and effects after two months. Dev Cogn Neurosci. 2012; 2: S192–204. https://doi.org/10.1016/j.dcn.2011.09.004 PMID: 22682908

39. Shelton JT, Elliott EM, Matthews RA, Hill BD, WmD G. The Relationships of Working Memory, Secondary Memory, and General Fluid Intelligence: Working Memory is Special. J Exp Psychol Learn Mem Cogn. 2010; 36:813–20. https://doi.org/10.1037/a0019046 PMID: 20438278

40. Conners C. Conners' Continuous Performance Test II: computer program for Windows technical guide and software manual. North Tonwanda (New York): Multi-Health Systems; 2000.

41. Ministry of Public Works. Atlas of urban vulnerability in Spain: methodology and contents. Madrid: Ministry of Public Works. Spanish Government; 2012. Available from: https://www.yumpu.com/en/document/view/28042109/atlas-of-urban-vulnerability-in-spain-methodology-and-contents.

42. de Onis M, Garza C, Onyango AW, Rolland-Cachera M-F. le Comité de nutrition de la Société française de pédiatrie. WHO growth standards for infants and young children. Arch Pediatr Organe Off Soc Francaise Pediatr. 2009; 16:47–53. https://doi.org/10.1016/j.arcped.2008.10.010 PMID: 19036567

43. Programme for International Student Assessment (PISA). In: OECD Programme for International Student Assessment [Internet]. 2012 [cited 2021 Oct7]. Available from: https://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm.

44. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001; 40:1337–45. https://doi.org/10.1097/00004583-200111000-00015 PMID: 11699809

45. Amato F, Rivas I, Viana M, Moreno T, Bouso L, Reche C, et al. Sources of indoor and outdoor PM2.5 concentrations in primary schools. Sci Total Environ. 2014; 490:757–65. https://doi.org/10.1016/j.scitotenv.2014.05.051 PMID: 24907610

46. Reche C, Viana M, Rivas I, Bouso L, Àlvarez-Pedrerol M, Alastuey A, et al. Outdoor and indoor UFP in primary schools across Barcelona. Sci Total Environ. 2014; 493:943–53. https://doi.org/10.1016/j.scitotenv.2014.06.072 PMID: 25003584

47. Rivas I, Viana M, Moreno T, Pandolfi M, Amato F, Reche C, et al. Child exposure to indoor and outdoor air pollutants in schools in Barcelona. Spain Environ Int. 2014; 69:200–12. https://doi.org/10.1016/j.envint.2014.04.009 PMID: 24875803

48. World Health Organization. Guidelines for Community Noise. Geneva, Switzerland; 1999. Available from: http://www.who.int/docstore/peh/noise/guidelines2.html.

49. Barcelona City Council. Statistical Yearbook of Barcelona City. Year 2013. Barcelona; 2013.

50. Basagaña X, Esnaola M, Rivas I, Amato F, Alvarez-Pedrerol M, Forns J, et al. Neurodevelopmental Deceleration by Urban Fine Particles from Different Emission Sources: A Longitudinal Observational Study. Environ Health Perspect. 2016; 124:1630–6. https://doi.org/10.1289/EHP209 PMID: 27128166

51. Forns J, Dadvand P, Esnaola M, Alvarez-Pedrerol M, López-Vicente M, Garcia-Esteban R, et al. Longitudinal association between air pollution exposure at school and cognitive development in school children over a period of 3.5 years. Environ Res. 2017; 159:416–21. https://doi.org/10.1016/j.envres.2017.08.031 PMID: 28858754

Bibliography

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Nature Precedings*, pages 1–1.

Anderson, P. (2002). Assessment and development of executive function (ef) during childhood. *Child Neuropsychology*, 8(2):7182. PMID: 12638061.

Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):148.

Chickering, D. M., Heckerman, D., and Meek, C. (2004). Large-sample learning of bayesian networks is np-hard. *J. Mach. Learn. Res.*, 5:12871330.

Cooper, G. F. and Herskovits, E. (1992). A bayesian method for the induction of probabilistic networks from data. *Mach. Learn.*, 9(4):309347.

Cornfield, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. In *Fed Proc*, volume 21, page 5861.

Dawber, T. R., Moore, F. E., and Mann, G. V. (1957). Ii. coronary heart disease in the framingham study. *American Journal of Public Health and the Nations Health*, 47(4_Pt_2):424. PMID: 13411327.

Diez-Roux, A. V. (1998). Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *American journal of public health*, 88(2):216222.

Esnaola, M. (2013). Biological structure discovery using bayesian networks. *Unpublished*.

Friedman, N. and Koller, D. (2000). Being bayesian about network structure. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, UAI'00, page 201210, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Gamez, J. A., Mateo, J. L., and Puerta, J. M. (2011). Learning bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Min. Knowl. Discov.*, 22(1-2):106148.

Grandjean, P. and Landrigan, P. J. (2014). Neurobehavioural effects of developmental toxicity. *The lancet neurology*, 13(3):330338.

Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, page 3748.

Guxens, M., Aguilera, I., Ballester, F., Estarlich, M., Fernandez-Somoano, A., Lertxundi, A., Lertxundi, N., Mendez, M. A., Tardon, A., Vrijheid, M., and Sunyer, J. (2012). Prenatal exposure to residential air pollution and infant mental development: Modulation by antioxidants and detoxification factors. *Environmental health perspectives*, 120(1):144149.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97109.

Hernn, M. A., Hernndez-Daz, S., Werler, M. M., and Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *American journal of epidemiology*, 155(2):176184.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *STATISTICAL SCIENCE*, 14(4):382417.

Koivisto, M. (2006). Advances in exact bayesian structure discovery in bayesian networks. In *Proceedings of the Twenty-Second Conference on*

Uncertainty in Artificial Intelligence, UAI'06, page 241248, Arlington, Virginia, USA. AUAI Press.

Kokonendji, C. C., Simplice, D.-G., and Clarice G. B., D. (2004). Some discrete exponential dispersion models: Poisson-tweedie and hinde-demtrio classes. *SORT-Statistics and Operations Research Transactions*, 28(2):201–214.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):1–21.

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, page 1517.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer, New York, NY [u.a.].

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43(7):e47e47.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, 26(1):139–140.

Rothman, K., Greenland, S., and Lash, T. (2008). *Modern Epidemiology*. Wolters Kluwer Health/Lippincott Williams & Wilkins.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., and Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, 42(8):10291040.

Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of The National Academy of Sciences of The United States Of America*, 74(12):54635467.

Szklo, M. and Nieto, J. (2007). *Epidemiology: Beyond the Basics*. Jones & Bartlett Learning.

Tang, M., Sun, J., Shimizu, K., and Kadota, K. (2015). Evaluation of methods for differential expression analysis on multi-group rna-seq count data. *BMC bioinformatics*, 16(1):1–14.

United-States (1964). *Smoking and health : report of the Advisory Committee to the Surgeon General of the Public Health Service.* Public Health Service publication ; no. 1103. U.S. Department of Health, Education, and Welfare, Public Health Service, Washington, D.C.

Wang, M., Beelen, R., Basagana, X., Becker, T., Cesaroni, G., de Hoogh, K., Dedele, A., Declercq, C., Dimakopoulou, K., Eeftens, M., et al. (2013). Evaluation of land use regression models for no2 and particulate matter in 20 european study areas: the escape project. *Environmental science & technology*, 47(9):43574364.

Wang, Z., Gerstein, M., and Snyder, M. (2009). Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1):5763.