

LETTER TO THE EDITOR

WILEY

Nonlinear effects of environment on childhood asthma susceptibility

To the Editor,

Environmental exposures such as air pollution, farm animals, and tobacco smoke have been established as risk factors for childhood asthma.¹ Single nucleotide polymorphisms (SNPs) associated with childhood asthma have also been identified in both candidate gene

and genome-wide association studies.^{2,3} However, the majority of studies consider genetic and environmental factors independently and assume the degree of risk conferred is uniform across all individuals.

As the impact of environmental risk factors may be influenced by an individual's genetic predisposition to the disease, a

TABLE 1 Association between parental smoking exposure in the first year of a child's life and the development of childhood asthma

Risk of childhood asthma							
		Unadjusted model ^a	Unadjusted for interaction effect ^b	Adjusted for interaction effect ^c	Low PRS tertile ^b	Middle PRS tertile ^b	High PRS tertile ^b
IOWBC	N	835	797	797	266	265	266
	OR (95% CI)	1.37 (0.93, 2.01)	1.39 (0.94, 2.07)	Parental smoking: 1.27 (0.84, 1.91) Interaction: 1.43 (0.77, 2.65)	0.92 (0.40, 2.14)	1.18 (0.57, 2.44)	1.76 (0.98, 3.17)
	p-Value	.119	.10	Parental smoking: .260 Interaction: .26	.85	.66	.06
MAAS	N	765	765	765	255	255	255
	OR (95% CI)	0.95 (0.59, 1.51)	0.94 (0.58, 1.51)	Parental smoking: 2.01 (1.07–3.79) Interaction: 0.30 (0.15, 0.62)	1.64 (0.64, 4.21)	1.53 (0.71, 3.31)	0.38 (0.15, 0.97)
	p-Value	.82	.79	Parental smoking: .03 Interaction: .001	.30	.27	.04
ALSPAC	N	3,334	2,802	2,802	935	938	939
	OR (95% CI)	1.15 (0.95, 1.40)	1.27 (1.01, 1.59)	Parental smoking: 1.26 (0.98, 1.61) Interaction: 0.91 (0.64, 1.29)	1.15 (0.72, 1.83)	1.27 (0.85, 1.88)	1.36 (0.96, 1.93)
	p-Value	.16	.04	Parental smoking: .07 Interaction: .58	.56	.24	.09

Note: A *p*-value threshold of .05 was used to evaluate whether there was weak or strong evidence for an association, values < 0.05 are in bold. Individuals were assigned into low, middle, and high PRS tertiles based on the distribution of the PRS in each cohort—IOWBC: low: (−2.17, −0.35), middle: (−0.36, 0.27), high: (0.28, 2.03); MAAS: low: (−1.83, 0.19), middle: (0.20, 0.79), high: (0.80, 3.12); ALSPAC: low: (−2.98, −0.70); middle: (−0.71, −0.13); high: (−0.14, 1.86). Covariates are gender, maternal asthma, paternal asthma, and sibling asthma.

Abbreviation: OR, odds ratio.

^aModel unadjusted for covariates or PRS-by-parental smoking interaction: asthma ~ parental smoking.

^bModel unadjusted for a PRS-by-parental smoking interaction: asthma ~ parental smoking + covariates.

^cModel adjusted for covariates and a PRS-by-parental smoking interaction: asthma ~ parental smoking + PRS (continuous variable) + PRS-by-parental smoking interaction + covariates.

number of studies have investigated the presence of possible gene-environment interactions.^{4,5} However, the majority of these studies have focused on specific candidate genes. Although genome-wide interaction studies have been performed, these are also SNP-wise analyses where each genotyped SNP is independently assessed against the environmental exposure and an SNP-by-exposure interaction term.⁶

Genetic variants identified in genome-wide association studies are often common variants that only confer small effects.⁶ As such, the power to detect interactions between exposures and individual SNPs is limited. Polygenic risk scores (PRSs) have been used as a means of combining the small effects of multiple SNPs across the genome in order to derive a better estimate of an individual's overall genetic risk.⁶ It is possible that PRSs may offer increased power to identify interaction effects. For example, Aschard et al.⁷ demonstrated that whilst no interactions for pulmonary function were identified among 26 independent SNP-by-smoking interaction analyses, combining the 26 SNPs into a PRS and conducting a PRS-by-smoking interaction analysis were able to uncover an interaction effect. Yet, interactions between environmental exposures and an individual's complete genetic profile have seldom been explored,^{8,9} with no such studies currently related to childhood asthma.

Arshad et al.¹⁰ previously identified a number of early life risk factors, such as parental smoking in the first year of a child's life, to be associated with childhood asthma development at age 10 in the Isle of Wight Birth Cohort (IOWBC). In this study, using the environmental exposure of parental smoking in the first year of a child's life as an exemplar, a PRS-by-parental smoking interaction meta-analysis was performed to: (i) evaluate the presence of an interaction between parental smoking and polygenic risk on the development of childhood asthma and (ii) assess whether the risk of parental smoking on the development of childhood asthma is in fact uniform among individuals of different genetic risk.

Data from three UK birth cohorts were used in this study: the IOWBC, Manchester Asthma and Allergy Study (MAAS), and The Avon Longitudinal Study of Parents and Children (ALSPAC) (see Supplementary Methods for cohort details, Table S1). Childhood asthma, defined as a combination of a doctor's diagnosis ever and the presence of wheeze or the use of asthma medication in the last 12 months, was evaluated at age 10 in the IOWBC ($n = 1368$ (201 cases)) and at age 11 in MAAS ($n = 898$ (116 cases)) and ALSPAC ($n = 4704$ (728 cases)). Based on a previously published list of independent asthma SNPs summarized across existing GWASs published until 2019, weighted 105-SNP PRSs were calculated for each individual with genotype data in each cohort using PRSice (see Supplementary Methods). Logistic regression analysis was used to assess the association of parental smoking and childhood asthma in each cohort (in the first year of life in the IOWBC and ALSPAC and at the antenatal follow-up in MAAS). The analysis was then repeated with the addition of an interaction term between the parental smoking and the PRS as a continuous variable. Next, individuals in each cohort were stratified into tertiles based on the cohort PRS to group individuals with a low, middle, or high genetic predisposition for childhood asthma.

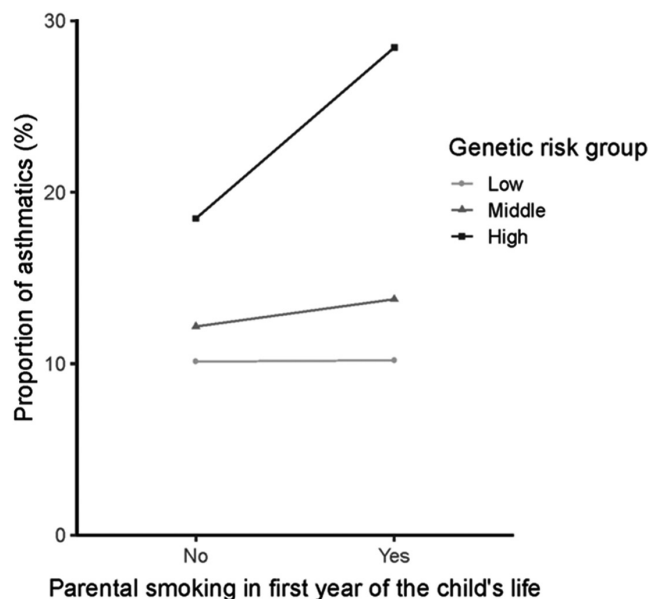


FIGURE 1 Association between parental smoking in the first year of life and childhood asthma in the IOWBC, stratified by PRS

The initial logistic regression analysis (without the interaction term) was then repeated in each genetic (PRS) subgroup. All analyses in the IOWBC (and where data were available in MAAS and ALSPAC at their respective time points, Table S1) were adjusted for gender, maternal asthma, paternal asthma, and sibling asthma. As the familial history of asthma may also account for some of the genetic effects being investigated, all analyses were also performed adjusting for gender only (see Supplementary Material). Finally, the stratified logistic regression analysis was meta-analyzed across all three cohorts using both fixed- and random-effect models using the R package "meta."

In the IOWBC, MAAS, and ALSPAC, data for asthma, parental smoking in the first years of life, and PRS were available for 835, 765, and 3,334 individuals, respectively (Table 1 and S2). In ALSPAC, there was some evidence for a 27% increased risk of developing childhood asthma when parents smoked in the first year of life, OR (95% CI): 1.27 (1.01, 1.59) but not in the IOWBC or MAAS ($p = .10$ and $p = .79$, respectively) (Table 1).

The interaction analysis showed no evidence for an interaction effect between the PRS and parental smoking in the development of childhood asthma; however, the stratified analysis did reveal a trend effect. In the IOWBC and ALSPAC, parental smoking was shown to confer a larger effect (with a higher degree of statistical evidence) among individuals with higher PRSs (Table 1, Figure 1). Strong evidence for this trend effect was demonstrated in the fixed-effect meta-analysis model ($p = .02$) (Figure 2). Similar meta-analysis results are presented when adjusting only for gender as a covariate (Table S3, Figure S1).

Particularly as childhood asthma is a highly heterogeneous disease, this study illustrates the importance of not evaluating environmental risk factors in isolation, as these too may be heterogeneous themselves. For example, this study demonstrates the risk conferred by parental smoking is not uniform but can vary depending on an individual's genetic susceptibility.

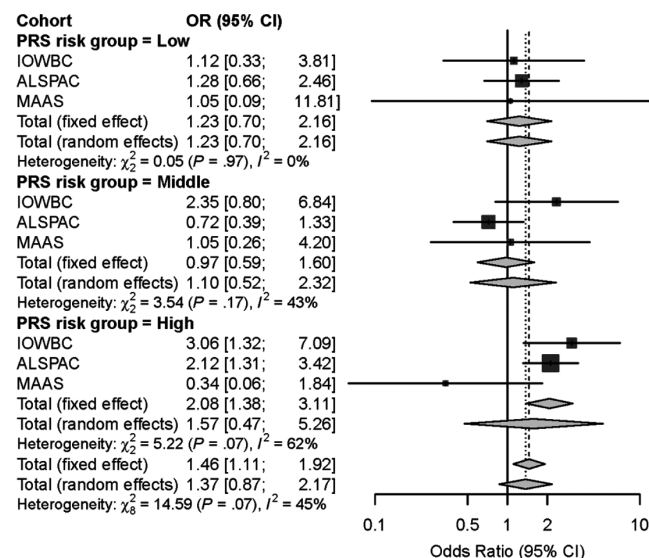


FIGURE 2 Forest plot showing the association of parental smoking in the first year of a child's life on the development of childhood asthma. The forest plot was constructed using summary statistics derived from the logistic regression analysis conducted across each PRS tertile in each cohort. The risk (odds ratio (OR) of parental smoking on the development of childhood asthma for each cohort (grey squares) with 95% confidence intervals (95% CI, from the solid black horizontal lines) is shown for each cohort, stratified by the PRS tertiles. The size of the grey squares corresponds to the cohort sample size. Subgroups with an odds ratio to the right (left) of the reference line (OR = 1, solid black vertical line) indicate an increased (decreased) risk of asthma among children exposed to parental smoking in the first year of life. Average effect sizes pooled across all three cohorts are presented for each PRS tertile and across all PRS tertiles, for both the fixed- and random-effect meta-analyses (diamonds). Dashed and dotted black vertical lines show the pooled effects across all PRS tertiles for the fixed- and random-effect models, respectively. In this analysis, covariates adjusted for included gender, maternal asthma, paternal asthma, and sibling asthma

However, it is important to acknowledge that this study was unable to offer conclusive evidence for a genetics-by-environment interaction (or direction of an interaction effect) based on the use-case of parental smoking in the first year of life. This may merely be a product of the inherent complexity of PRS-environment effects being explored or the need to account for additional complex influences such as collider bias.¹¹

The conflicting findings in MAAS in comparison to ALSPAC and IOWBC may stem from antenatal parental smoking rather than parental smoking within the first year of the child's life (as in the IOWBC and ALSPAC) being evaluated in MAAS. However, this may also be explained by inherent cohort differences; MAAS conducted selective recruitment from specific regions of South Manchester and Cheshire (see Supplementary Methods) whilst the IOWBC and ALSPAC are unselected cohorts representative of the general population. Although there was a 91% agreement between parental smoking status at recruitment and age three in MAAS (time points at which parental smoking data were available) and it is likely that parental smoking status reported at recruitment

was therefore representative of parental smoking in the first year of a child's life, it is highly possible that parental smoking behavior may have been different in the year immediately following the child's birth.

Ultimately, further investigation is needed to untangle the association of established environmental risk factors for asthma and improve the accuracy of their main effects. However, this study highlights the importance of future studies considering heritable (genetic) risk and any potential interaction effects that may influence the magnitude of risk estimates. Such studies would be particularly important for the formulation of public health policies and guidelines. Similarly, if associations with risk factors are only observed among certain subgroups of individuals, developing personalized prediction models using only predictors relevant within those specific patient subgroups could improve upon the limited performance of current asthma prediction models.

CONTENT

This publication is the work of the authors and Dilini M. Kothalawala & John W. Holloway will serve as guarantors for the contents of this paper.

KEYWORDS

ALSPAC, childhood asthma, meta-analysis, parental smoking, polygenic risk score, PRS-by-environment interactions

FUNDING INFORMATION

This work was supported by the NIHR Southampton Biomedical Research Centre and a University of Southampton Presidential Research Studentship. The 10-year follow-up of the IOWBC was supported by the National Asthma Campaign, UK (Grant No 364). MAAS was supported by the Asthma UK Grants No 301 (1995-1998), No 362 (1998-2001), No 01/012 (2001-2004), No 04/014 (2004-2007), BMA James Trust (2005), and the JP Moulton Charitable Foundation (2004-current), the North West Lung Centre Charity (1997-current), and the Medical Research Council (MRC) G0601361 (2007-2012), MR/K002449/1 (2013-2014), and MR/L012693/1 (2014-2018). UNICORN (Unified Cohorts Research Network): Disaggregating asthma MR/S025340/1. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. GWAS data were generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. RG is in part funded through Wellcome Trust Strategic Award 108818/15/Z.

AUTHOR CONTRIBUTIONS

Dilini M. Kothalawala: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (lead); Project administration (lead); Software (equal); Validation (equal); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Veronique B.N. Weiss:** Formal analysis (equal); Investigation (equal); Software (equal); Validation (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Latha Kadalayil:** Methodology (supporting); Writing – review & editing (supporting).

Raquel Granell: Validation (equal); Writing – review & editing (supporting). **John Curtin:** Validation (supporting); Writing – review & editing (supporting). **Clare Suzanne Murray:** Resources (equal); Writing – review & editing (supporting). **Angela Simpson:** Funding acquisition (equal); Resources (equal); Writing – review & editing (supporting). **Adnan Custovic:** Funding acquisition (equal); Resources (equal); Writing – review & editing (supporting). **William J. Tapper:** Methodology (supporting); Supervision (equal); Writing – review & editing (supporting). **Faisal I. Rezwan:** Supervision (equal); Writing – review & editing (supporting). **Hasan Arshad:** Funding acquisition (equal); Methodology (supporting); Resources (equal); Supervision (equal); Writing – review & editing (supporting). **John W. Holloway:** Conceptualization (lead); Funding acquisition (equal); Methodology (lead); Project administration (supporting); Resources (equal); Supervision (equal); Writing – review & editing (supporting).

Dilini M. Kothalawala^{1,2} 
 Veronique B. N. Weiss¹ 
 Latha Kadalayil^{1,3} 
 Raquel Granell⁴ 
 John A. Curtin^{5,6} 
 Clare S. Murray⁵ 
 Angela Simpson⁵ 
 Adnan Custovic⁷ 
 William J. Tapper¹ 
 Faisal I. Rezwan^{1,8} 
 Syed Hasan Arshad^{1,3,9} 
 John W. Holloway^{1,2} 

STELAR/UNICORN investigators

¹Human Development and Health, Faculty of Medicine,
University of Southampton, Southampton, UK

²NIHR Southampton Biomedical Research Centre, University
Hospital Southampton, Southampton, UK

³Clinical and Experimental Sciences, Faculty of Medicine,
University of Southampton, Southampton, UK

⁴MRC Integrative Epidemiology Unit, Bristol Medical School,
University of Bristol, Bristol, UK

⁵Division of Infection, Immunity, and Respiratory Medicine,
Manchester Academic Health Science Centre, School of
Biological Sciences, Manchester University Hospital NHS
Foundation Trust, University of Manchester, Manchester, UK

⁶NIHR Manchester Biomedical Research Unit, Manchester, UK

⁷Imperial College of Science, Technology, and Medicine, National
Heart and Lung Institute, London, UK

⁸Department of Computer Science, Aberystwyth University,
Aberystwyth, UK

⁹The David Hide Asthma and Allergy Research Centre, St.
Mary's Hospital, Newport, Isle of Wight, UK

Editor: Ömer Kalayci

Correspondence

John W. Holloway, Human Development and Health, Faculty
of Medicine, University of Southampton, Southampton, UK.

Email: j.w.holloway@soton.ac.uk

ORCID

Dilini M. Kothalawala  <https://orcid.org/0000-0002-5804-0457>
 Veronique B. N. Weiss  <https://orcid.org/0000-0003-4116-950X>
 Latha Kadalayil  <https://orcid.org/0000-0002-3757-5487>
 Raquel Granell  <https://orcid.org/0000-0002-4890-4012>
 John A. Curtin  <https://orcid.org/0000-0001-6299-9110>
 Clare S. Murray  <https://orcid.org/0000-0002-8961-8055>
 Angela Simpson  <https://orcid.org/0000-0003-2733-6666>
 Adnan Custovic  <https://orcid.org/0000-0001-5218-7071>
 William J. Tapper  <https://orcid.org/0000-0002-5896-1889>
 Faisal I. Rezwan  <https://orcid.org/0000-0001-9921-222X>
 Syed Hasan Arshad  <https://orcid.org/0000-0001-5988-235X>
 John W. Holloway  <https://orcid.org/0000-0001-9998-0464>

REFERENCES

1. Subbarao P, Mandhane PJ, Sears MR. Asthma: epidemiology, etiology and risk factors. *CMAJ*. 2009;181(9):E181-E190.
2. Meyers DA, Bleecker ER, Holloway JW, Holgate ST. Asthma genetics and personalised medicine. *Lancet Respir Med*. 2014;2:405-415.
3. Morales E, Duffy D. Genetics and gene-environment interactions in childhood and adult onset asthma. *Front Pediatr*. 2019;7:499.
4. London SJ, Romieu I. Gene by environment interaction in asthma. *Annu Rev Public Health*. 2009;30:55-80.
5. Ege MJ, Strachan DP, Cookson WOCM, et al. Gene-environment interaction for childhood asthma and exposure to farming in Central Europe. *J Allergy Clin Immunol*. 2011;127:138-44.e4.
6. Hernandez-Pacheco N, Pino-Yanes M, Flores C. Genomic predictors of asthma phenotypes and treatment response. *Front Pediatr*. 2019;7:6.
7. Aschard H, Tobin MD, Hancock DB, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int J Epidemiol*. 2017;46(3):894-904.
8. Sikdar S, Wyss AB, Lee MK, et al. Interaction between genetic risk scores for reduced pulmonary function and smoking, asthma and endotoxin. *Thorax*. 2021;76(12):1219-1226.
9. Qi Q, Chu AY, Kang JH, et al. Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med*. 2012;367:1387-1396.
10. Arshad SH, Kurukulaaratchy RJ, Fenn M, Matthews S. Early life risk factors for current wheeze, asthma, and bronchial hyperresponsiveness at 10 years of age. *Chest*. 2005;127:502-508.
11. Akimova ET, Breen R, Brazel DM, Mills MC. Gene-environment dependencies lead to collider bias in models with polygenic scores. *Sci Rep*. 2021;11:9457.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Supplementary Material

Methods

Study populations

Isle of Wight Birth Cohort (IOWBC)

The IOWBC, also known as the second generation (F1) cohort, is based on the Isle of Wight, off the south coast of England, UK. It is a whole population, single-centre prospective cohort study established in 1989 to explore the natural history and development of asthma and other allergic diseases in early life¹. Between 1989 and 1990, 1509 women gave birth to 1536 children on the Isle of Wight. From these births, 1456 children were recruited into the IOWBC study and followed up from birth and at 1, 2, 4, 10, 18 and 26 years, with high retention rates of 94.0%, 84.5%, 83.7%, 94.3%, 90.2% and 70.9%, respectively¹. The IOWBC cohort comprises of individuals predominantly of Caucasian ethnicity (98%).

Study recruitment, participant details and available data have been previously described¹ and further information can be found at www.allergyresearch.org.uk/ or from: Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol*. 2018 Aug 1;47(4):1043-1044i. doi: 10.1093/ije/dyy023.

Ethical approval was obtained from the Isle of Wight Local Research Ethics Committee at recruitment and 1, 2 and 4-year assessments (No. 05/89) and 10-year assessments including genetic studies (No. 18/98). Prior to participation in the study at each follow-up, written informed consent was obtained from parents of children, and assent from children (where applicable).

Genotyping and Imputation

Blood samples were collected from the child at multiple time-points - from heel pricks at 7 days of age (collected on Guthrie cards) and in later childhood at 10, 18 and 26 years. DNA from the peripheral blood samples of 1067 individuals in the IOWBC were isolated and underwent genome-wide genotyping using the Illumina InfiniumOmni2.5-8v1.3 microarray. Standard quality control had been performed to exclude samples with low call rate (<97%) and SNPs with call rate <95%, minor allele frequency (MAF)<0.005 and deviation from Hardy-Weinberg equilibrium (HWE) (p-value <1x10⁻⁸) prior to imputation. Alleles had also been updated to match the direction (forward) and coordinates of reference dataset GRCh37². Data were pre-phased (EAGLE2)³ and imputed (PBWT)⁴ using the Sanger Imputation Services (Oxford, UK).

Further quality control was performed to prepare the imputed genotype data for analyses. This included the retention of data with an imputation quality >80%. SNPs were again filtered to exclude those with call rate <95%, MAF<0.01 and deviation from HWE (p-value <1x10⁻⁶). Samples were further filtered to remove: call-rate <97%; extreme heterozygosity ($\pm 3SD$ of the mean F-coefficient); gender mismatch and one individual of each related pair (3rd degree relations or closer, pi-hat ≥ 0.125). Population structure was assessed by principal component analysis (PCA), comparing the IOWBC with the European descent (CEU), Yoruba (YRI), Hans Chinese (CHB) and Japanese (JPT) HapMap3 reference populations. Non-European individuals were excluded based on a visual

inspection of the PCA plot. A final dataset of 977 individuals with genotype data for 7,236,427 SNPs was retained for downstream analyses.

In addition, five loss of function variants of the *Fillagrin* gene (R501X, 2282del, or S3247X, 3702delG and R2447X) were separately genotyped using GoldenGate Genotyping Assays (Illumina, Inc, SanDiego CA) on the BeadXpress Veracode platform per Illumina's protocol⁵. In brief, 1,248 blood samples (from 1,211 individuals and 37 replicates) were fragmented, hybridised to allele-specific primer sets and subject to extension/ligation reactions. Samples were then hybridised to the Veracode bead pool for processing by the BeadXpress reader. Allele determination was based on a GenCall score >0.25. Scores below this quality threshold were deemed "no calls".

For downstream analyses, quality controlled genotype data for the R501X variant (n=924) was added to the quality controlled imputed genome-wide genotype profiles detailed above, resulting in a final dataset of 924 genotype profiles consisting of 7,236,428 SNPs.

Manchester Asthma and Allergy Study (MAAS)

The Manchester Asthma and Allergy Study (MAAS) is an unselected birth cohort established to study the development of asthma and other atopic disorders in childhood⁶. Participants were recruited into the study from 50 square miles of South Manchester and Cheshire (within the maternity catchment area of the Wythenshawe and Stepping Hill Hospitals). Between 1995 and 1997, 1211 women (≤ 10 weeks pregnant) were recruited into the study and 1184 children were subsequently followed up at 1, 3, 5, 8, 11, 13-16 and 18 years. The MAAS cohort consists of a stable mixed urban-rural population (~89% Caucasian).

Study recruitment, participant details and available data have been previously described⁶ and further information can be found at www.allergyresearch.org.uk/.

Genotyping and Imputation

DNA samples were genotyped using the Illumina 610 quad chip. Genotypes were called using the Illumina GenCall application following the manufacturer's instructions. Prior to imputation samples were excluded on the basis of gender mismatches; minimal or excessive heterozygosity, genotyping call rates of <97%. SNPs were excluded if they had call rates of <95%, minor allele frequencies of <0.5% and HWE $p < 3 \times 10^{-8}$. Prior to imputation each chromosome was prephased using EAGLE2 (v2.0.5) as recommended by the Sanger Imputation Server. We then imputed with PBWT with the Haplotype Reference Consortium (release 1.1) of 32,470 reference genomes using the Sanger Imputation Server.

Imputed genotype data was further quality controlled to exclude SNPs with $INFO < 0.80$, $MAF < 0.01$ and deviations from HWE ($p\text{-value} \leq 1 \times 10^{-8}$). One individual from each pair of siblings or cryptic relations and non-Caucasians were also excluded. A final dataset 102 SNPs from 852 individuals were used to generate polygenic risk score in MAAS.

Avon Longitudinal Study of Parents and Children (ALSPAC)

The Avon Longitudinal Study of Parents and Children was established in Bristol (in the southwest of England), as one of a series of birth cohorts focused on investigating modifiable factors influencing health and development in Europe^{7,8}.

Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled is 14,541 (for these at least one questionnaire has been returned or a "Children in Focus" clinic had been attended by 19/07/99). Of these initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age.

When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected from the age of seven onwards (and potentially abstracted from obstetric notes) there are data available for more than the 14,541 pregnancies mentioned above. The number of new pregnancies not in the initial sample (known as Phase I enrolment) that are currently represented on the built files and reflecting enrolment status at the age of 24 is 913 (456, 262 and 195 recruited during Phases II, III and IV respectively), resulting in an additional 913 children being enrolled. The phases of enrolment are described in more detail in the cohort profile paper and its update. The total sample size for analyses using any data collected after the age of seven is therefore 15,454 pregnancies, resulting in 15,589 fetuses. Of these 14,901 were alive at 1 year of age.

A 10% sample of the ALSPAC cohort, known as the Children in Focus (CiF) group, attended clinics at the University of Bristol at various time intervals between 4 to 61 months of age. The CiF group were chosen at random from the last 6 months of ALSPAC births (1432 families attended at least one clinic). Excluded were those mothers who had moved out of the area or were lost to follow-up, and those partaking in another study of infant development in Avon.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool" and reference the following webpage:
<http://www.bristol.ac.uk/alspac/researchers/our-data/>

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

Genotyping and Imputation

DNA was collected from 11,343 individuals in ALSPAC and genome-wide genotype data was performed for 8365 children⁹. Data was genotyped using Illumina HumanHap550 quad chip genotyping platforms. Standard quality control excluded samples with low call rate (<97%), minimal or excessive heterozygosity, gender mismatch, insufficient sample replication (IBD<0.80) and cryptic relatedness (IBD>0.1). Multidimensional scaling analysis was used to assess population structure in ALSPAC against the European descent (CEU), Yoruba (YRI), Hans Chinese (CHB) and Japanese (JPT) Hapmap II (release 22) reference populations, and non-European individuals were excluded. SNPs with call rate <95%, MAF<0.01 and deviation from HWE (p-value <5x10⁻⁷) prior to imputation. Alleles

were also updated to match the direction (forward) and coordinates of reference dataset GRCh37². Genotypes were then imputed using IMPUTE version 2.2.2, with the 1000 genomes phase 1 version 3 reference panel.

Imputed genotype data was further quality controlled to exclude SNPs with INFO<0.80, MAF<0.01 and deviations from HWE (p-value $\leq 1 \times 10^{-8}$). A final dataset of 7,977 individuals with genotype profiles of 8,099,747 SNPs were retained for downstream analysis.

Polygenic risk score

To construct the PRS, 128 independent SNPs associated with asthma were considered. These SNPs were identified from a recent study conducted by El-Husseini et al. which provided an updated summary of independent SNPs associated with asthma from published GWASs between 2007 and 2019¹⁰. In brief, SNPs with genome-wide significance ($p < 3 \times 10^{-8}$) were identified and tested for independence in European populations using the LD Matrix tool on LD Link. SNPs were considered independent if the linkage disequilibrium (LD) correlation (r^2) was less than 0.05.

The list of 128 SNPs were summarised from a culmination of different asthma GWASs. Therefore, to construct the PRS specifically for childhood asthma, summary statistics for the 128 SNPs were extracted from a single GWAS study conducted by Ferreira et al. which used data from UK Biobank to identify SNPs associated with a childhood onset asthma phenotype¹¹. SNPs were included in the construction of the PRS if genotype data (post quality control) was available in the IOWBC and summary statistics were available in Ferreira et al.'s GWAS. Where data for a SNP was unavailable in either the IOWBC or within the GWAS summary statistics, the closest proxy SNP in high LD ($r^2 > 0.8$) within the European British in England and Scotland (GBR) population with data available was sourced using the LD proxy tool on LD Link¹².

The summary statistic data extracted for each SNP from the GWAS included: i) the SNP effect size - to weight each SNP in the PRS, and ii) the SNP p-value - to determine the inclusion of SNPs in the score using the thresholding method. Where proxy SNPs were used, the effect size and p-value of the original SNP were used.

A weighted childhood asthma PRS was constructed based on the clumping and thresholding method using PRSice¹³. The PRS was calculated as the sum of an individual's risk alleles weighted by the allele effect size as estimated from the GWAS study (Equation 1).

$$PRS = \sum_i^N X_i \beta_i$$

Equation 1 Formula for calculating a weighted polygenic risk score

The polygenic risk score (PRS) is calculated as the sum of all SNPs included in the score (N). For each SNP in the PRS (i), the dosage of the risk allele (X) is weighted by its GWAS effect size estimate (β).

By default, PRSice performs clumping and removes SNPs in high LD ($r^2 > 0.1$) within a 250kb window and calculates scores across all possible p-value thresholds. To select the best PRS from all the scores calculated from the thresholding method, PRSice uses Nagalkere's R^2 goodness of fit statistic which evaluates how well the score explains the variance in the binary phenotype. Alongside Nagalkere's R^2 , the score which offered the highest AUC across 2000 bootstrapped samples was considered.

In IOWBC, only SNPs for which data was available for all individuals were considered and the best PRS consisted of 105 SNPs. Whilst the same approach was used to construct the PRS in MAAS, SNPs for which data was available for at least 90% of individuals were included in ALSPAC (to calculate the PRS, information on the minor allele frequency of missing SNPs was used). Where proxy SNPs were used in the IOWBC, the corresponding original SNP detailed in the list of 128 independent asthma SNPs was first considered upon replication in MAAS and ALSPAC. Where SNPs were unavailable,

proxy SNPs were sourced as previously described. If a proxy SNP was unavailable, the SNP was excluded from the cohort's PRS. As a result, 102 and 97 of the 105 SNPs were available to calculate each individual's PRS in MAAS and ALSPAC, respectively.

Results

Table S1 Definitions of variables used for analysis in each cohort

Variable	Variable Definitions		
	IOWBC	MAAS	ALSPAC
Gender	Gender (male/female)	Gender (male/female)	Gender (male/female)
Maternal asthma	Maternal asthma ever or wheezing attacks (no/yes) self-reported at recruitment	Maternal asthma ever (no/yes) self-reported at recruitment	Maternal asthma ever (no/yes) self-reported at 32 weeks gestation
Paternal asthma	Paternal asthma or wheezing attacks (no/yes) self-reported at recruitment	Paternal asthma ever (no/yes) self-reported at recruitment	Paternal asthma/wheezing after child was born (no/yes) was mother-reported when the study child was 8 months old
Sibling asthma	Siblings with asthma (no/yes) reported by parents at recruitment	Not available	Not available
Parental smoking in the first year of life	Maternal or paternal smoking inside the house (no/yes) reported at the 1-year follow-up (when the child was approximately 9 months)	Maternal or paternal smoking (no/yes) reported at recruitment	Maternal or paternal smoking (no/yes) when child was 8 months
Chest infection in the first year of life	Recurrent chest infections (no/yes) - parent-reported at the 1-year follow-up	Wheeze with a cold (no/yes) – parent-reported at the 1-year follow-up (surrogate variable)	Wheeze caused by chest infection in the last 12 months (no/yes) - reported by the mother at 18 months
Chest infection in early life	Recurrent chest infections (no/yes) - parent-reported at the 2-year follow-up	Wheeze with a cold (no/yes) – parent-reported at the 3-year follow-up (surrogate variable)	Wheeze caused by chest infection in the last 12 months (no/yes) - reported by the mother at 30 months
Eczema	Child currently suffering from eczema (no/yes) – parent-reported at the 4-year follow-up	Child currently suffering from eczema (no/yes) – parent-reported at the 5-year follow-up	Eczema in the last 12 months (no/yes) - reported by the mother at 3.5 years
Atopy	Sensitisation (positive skin prick test) to one or more food/aeroallergens at the 4-	Sensitisation (positive skin prick test) to one or more food/aeroallergens at the 4-	Sensitisation (positive skin prick test) to one or more food/aeroallergens at 7.5

	year follow-up	year follow-up	years
PRS	Polygenic risk scores calculated based on a 105-SNP PRS developed in the IOWBC	Polygenic risk scores calculated based on a 105-SNP PRS developed in the IOWBC (102/105 SNPs available in MAAS)	Polygenic risk scores calculated based on a 105-SNP PRS developed in the IOWBC (97/105 SNPs available in ALSPAC)
Asthma	Doctor diagnosed asthma ever and self-reported wheezing in the last 12 months and/or use of asthma medication at the 10-year follow-up	Doctor diagnosed asthma ever and self-reported wheezing in the last 12 months and/or use of asthma medication at the 11-year follow-up	Doctor diagnosed asthma ever and self-reported wheezing/asthma or medication in the last 12 months at 128 months (10.7 years)

Table S2 Descriptive statistics of variables included in the meta-analysis for each cohort

Variables [†]	Number of individuals (% asthmatic)		
	IOWBC N=835 (14.9)	MAAS N=765 (12.8)	ALSPAC N=3,334 (15.4)
Gender			
Male	426 (51.0%)	353 (46.1%)	1652 (49.6%)
Female	409 (49.0%)	352 (46.0%)	1682 (50.5%)
Maternal asthma			
No	738 (88.4%)	616 (80.5%)	1715 (53.1%)
Yes	92 (11.0%)	149 (19.5%)	1518 (47.0%)
Paternal asthma			
No	744 (89.1%)	656 (85.8%)	1651 (58.1%)
Yes	82 (9.8%)	109 (14.2%)	1192 (42.0%)
Sibling asthma			
No	553 (66.2%)	NA	NA
Yes	253 (30.3%)	NA	NA
PRS			
Asthmatics	-0.10 (-2.17, 2.03)	0.45 (-1.83, 2.46)	-0.21 (-2.29, 1.93)
Non-asthmatic	0.17 (-1.57, 1.49)	0.77 (-1.30, 3.12)	-0.45 (-2.98, 1.86)

Descriptive statistics for variables included in the interaction analysis in each cohort are presented as the number of individuals in each category, with the proportion of asthmatic individuals reported in brackets (%). For the continuous variable, PRS, the mean score is given for both asthmatics and non-asthmatic individuals, with the PRS range (min, max) in brackets.

[†] Variable definitions and time points for data collection in each cohort are detailed in Table S1.

Table S3 Association between parental smoking exposure in the first year of a child's life and the development of childhood asthma

Risk of childhood asthma							
		Unadjusted model [†]	Unadjusted for interaction effect [‡]	Adjusted for interaction effect [§]	Low PRS tertile [‡]	Middle PRS tertile [‡]	High PRS tertile [‡]
IOWBC	N	835	835	835	279	278	278
	OR	1.37	1.34	Parental smoking: 1.23 (0.82, 1.84)	0.93	1.07	1.74
	(95% CI)	(0.93, 2.01)	(0.91, 1.98)	Interaction: 1.47 (0.80, 2.69)	(0.41, 2.17)	(0.52, 2.17)	(0.98, 3.08)
	P-value	0.11	0.10	Parental smoking: 0.32 Interaction: 0.21	0.89	0.88	0.06
MAAS	N	765	765	765	255	255	255
	OR	0.95	0.95	Parental smoking: 1.93 (1.03-3.60)	1.55	1.51	0.42
	(95% CI)	(0.59, 1.51)	(0.59, 1.51)	Interaction: 0.33 (0.16, 0.68)	(0.62, 3.93)	(0.71, 3.21)	(0.17, 1.05)
	P-value	0.82	0.81	Parental smoking: 0.03 Interaction: 0.002	0.35	0.28	0.06
ALSPAC	N	3,334	3,334	2,802	935	938	939
	OR	1.15	1.15 (0.95, 1.40)	Parental smoking: 1.19 (0.96, 1.47)	0.94 (0.63, 1.41)	1.27 (0.90, 1.80)	1.23 (0.91, 1.67)
	(95% CI)	(0.95, 1.40)		Interaction: 1.03 (0.75, 1.40)			
	P-value	0.16	0.16	Parental smoking: 0.12 Interaction: 0.86	0.76	0.18	0.18

OR: odds ratio

A p-value threshold of 0.05 was used to evaluate if there was weak or strong evidence for an association.

Individuals were assigned into low, middle and high PRS tertiles based on the distribution of the PRS in each cohort – IOW= low: (-2.17, -0.35), middle: (-0.36, 0.27), high: (0.28- 2.03); MAAS= low: (-1.83, 0.19), middle: (0.20, 0.79), high: (0.80, 3.12) ; ALSPAC: low: (-2.97, -0.70); middle: (-0.71, -0.13); high: (-0.14, 1.86).

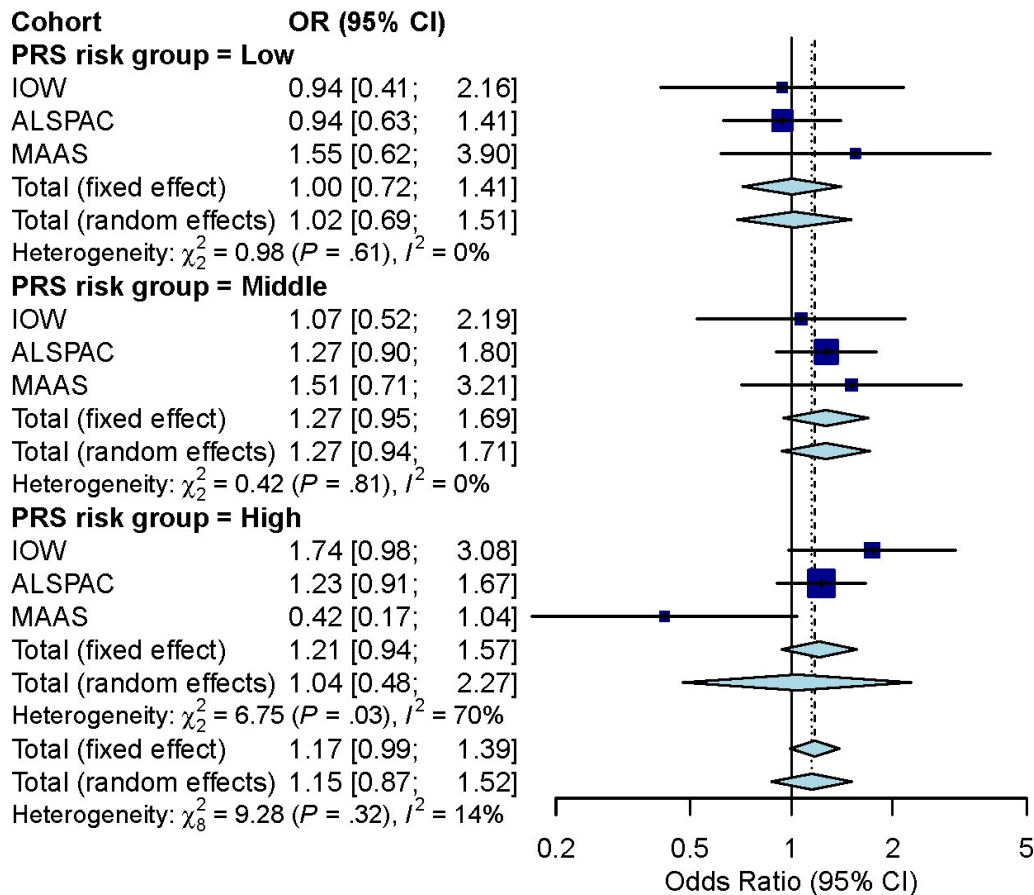
[†] Model unadjusted for covariates or PRS-by-parental smoking interaction: asthma ~ parental smoking

[‡] Model unadjusted for a PRS-by-parental smoking interaction: asthma ~ parental smoking + covariates

[§] Model adjusted for a PRS-by-parental smoking interaction: asthma ~ parental smoking + PRS (continuous variable) + PRS-by-parental smoking interaction + covariates.

Covariates adjusted for included: gender.

Figure S1 Forest plot showing the association of parental smoking in the first year of a child's life on the development of childhood asthma



The forest plot was constructed using summary statistics derived from the logistic regression analysis conducted across each PRS tertile in each cohort. The risk (odds ratio (OR) of parental smoking on the development of childhood asthma for each cohort (grey squares) with 95% confidence intervals (95% CI, from the solid black horizontal lines) is shown for each cohort, stratified by the PRS tertiles. The size of the grey squares corresponds to the cohort sample size. Subgroups with an odds ratio to the right (left) of the reference line (OR=1, solid black vertical line) indicate an increased (decreased) risk of asthma among children exposed to parental smoking in the first year of life. Average effect sizes pooled across all three cohorts are presented for each of the three PRS tertiles, for both the fixed and random effect meta-analyses (diamonds). Dashed and dotted black vertical lines show the pooled effects across all PRS tertiles for the fixed and random effect models, respectively. In this analysis, covariates adjusted for included: gender.

References

1. Arshad SH, Holloway JW, Karmaus W, et al. Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *International Journal of Epidemiology* 2018;47:1043-4i.
2. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 2016;48:1279-83.
3. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* 2016;48:1443-8.
4. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* 2014;30:1266-72.
5. Ziyab AH, Karmaus W, Yousefi M, et al. Interplay of filaggrin loss-of-function variants, allergic sensitization, and eczema in a longitudinal study covering infancy to 18 years of age. *PLoS One* 2012;7:e32721.
6. Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatric Allergy and Immunology* 2002;13:32-7.
7. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* 2013;42:111-27.
8. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology* 2012;42:97-110.
9. Taylor AE, Jones HJ, Sallis H, et al. Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* 2018;47:1207-16.
10. El-Husseini ZW, Gosens R, Dekker F, Koppelman GH. The genetics of asthma and the promise of genomics-guided drug target discovery. *The Lancet Respiratory Medicine* 2020;8:1045-56.
11. Ferreira MAR, Mathur R, Vonk JM, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *American Journal of Human Genetics* 2019;104:665-84.
12. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Frontiers in Genetics* 2020;11:157.
13. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols* 2020;15:2759-72.