2021-05-10

Environmental DNA reveals links between abundance and composition of airborne grass pollen and respiratory health

Rowney, FM

http://hdl.handle.net/10026.1/17880

Current Biology

Environmental DNA reveals links between abundance and composition of airborne grass pollen and respiratory health

Highlights

- Airborne grass pollen assemblages are quantitatively structured in space and time

- The respiratory health impacts of grass pollen may vary according to grass species

Authors

Francis M. Rowney,
Georgina L. Brennan,
Carsten A. Skjøth, ...,
Benedict Wheeler,
Nicholas J. Osborne, Simon Creer

Correspondence

f.rowney@exeter.ac.uk (F.M.R.),
g.l.b.doonan@gmail.com (G.L.B.),
b.w.wheeler@exeter.ac.uk (B.W.),
n.osborne@uq.edu.au (N.J.O.),
s.creer@bangor.ac.uk (S.C.)

In brief

Rowney and Brennan et al. explore relationships between grass pollen and human health using a nationwide, ecological approach to biomonitoring. Using eDNA and qPCR, we provide the first evidence that certain grass species may have disproportionate impacts on respiratory health (accounting for other factors, e.g., overall grass pollen concentrations).

CellPress

Report

Environmental DNA reveals links between abundance and composition of airborne grass pollen and respiratory health

Francis M. Rowney,[1,2,13,*] Georgina L. Brennan,[3,4,13,14,*] Carsten A. Skjøth,[5] Gareth W. Griffith,[6] Rachel N. McInnes,[7] Yolanda Clewlow,[7] Beverley Adams-Groom,[5] Adam Barber,[7] Natasha de Vere,[6,8] Theo Economou,[7,9] Matthew Hegarty,[6] Helen M. Hanlon,[7] Laura Jones,[8] Alexander Kurganskiy,[5,10] Geoffrey M. Petch,[5] Caitlin Potter,[6] Abdullah M. Rafiq,[3] Amena Warner,[11] The PollerGEN Consortium, Benedict Wheeler,[1,*] Nicholas J. Osborne,[1,12,*] and Simon Creer[3,*]

[1]European Centre for Environment and Human Health, University of Exeter, Knowledge Spa, Royal Cornwall Hospital, Truro TR1 3HD, UK
[2]School of Geography, Earth and Environmental Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK
[3]School of Natural Sciences, Bangor University, Deiniol Road, Bangor LL57 2UW, UK
[4]Centre for Environmental and Climate Science/Aquatic Ecology, Department of Biology, Lund University, 223 62 Lund, Sweden
[5]School of Science and the Environment, University of Worcester, Worcester WR2 6AJ, UK
[6]IBERS, Aberystwyth University, Aberystwyth SY23 3FL, UK
[7]Met Office, Fitzroy Road, Exeter EX1 3PB, UK
[8]National Botanic Garden of Wales, Llanarthne SA32 8HN, UK
[9]Department of Mathematics, University of Exeter, North Park Road, Exeter EX4 4QF, UK
[10]Department of Geography, University of Exeter, Penryn Campus, Treliever Road, Penryn TR10 9FE, UK
[11]Allergy UK, Edgington Way, Sidcup DA14 5BH, UK
[12]School of Public Health, The University of Queensland, Herston Road, Brisbane, QLD 4006, Australia
[13]These authors contributed equally
[14]Lead contact
*Correspondence: f.rowney@exeter.ac.uk (F.M.R.), g.l.b.doonan@gmail.com (G.L.B.), b.w.wheeler@exeter.ac.uk (B.W.), n.osborne@uq.edu.au (N.J.O.), s.creer@bangor.ac.uk (S.C.)
https://doi.org/10.1016/j.cub.2021.02.019

SUMMARY

Grass (Poaceae) pollen is the most important outdoor aeroallergen,[1] exacerbating a range of respiratory conditions, including allergic asthma and rhinitis ("hay fever").[2–5] Understanding the relationships between respiratory diseases and airborne grass pollen with a view to improving forecasting has broad public health and socioeconomic relevance. It is estimated that there are over 400 million people with allergic rhinitis[6] and over 300 million with asthma, globally,[7] often comorbidly.[8] In the UK, allergic asthma has an annual cost of around US$ 2.8 billion (2017).[9] The relative contributions of the >11,000 (worldwide) grass species (C. Osborne et al., 2011, Botany Conference, abstract) to respiratory health have been unresolved,[10] as grass pollen cannot be readily discriminated using standard microscopy.[11] Instead, here we used novel environmental DNA (eDNA) sampling and qPCR[12–15] to measure the relative abundances of airborne pollen from common grass species during two grass pollen seasons (2016 and 2017) across the UK. We quantitatively demonstrate discrete spatiotemporal patterns in airborne grass pollen assemblages. Using a series of generalized additive models (GAMs), we explore the relationship between the incidences of airborne pollen and severe asthma exacerbations (sub-weekly) and prescribing rates of drugs for respiratory allergies (monthly). Our results indicate that a subset of grass species may have disproportionate influence on these population-scale respiratory health responses during peak grass pollen concentrations. The work demonstrates the need for sensitive and detailed biomonitoring of harmful aeroallergens in order to investigate and mitigate their impacts on human health.

RESULTS AND DISCUSSION

The quantitative distribution of pollen from grass species is structured in space and time

Using large-scale sampling and qPCR, we show that grass pollen species assemblages vary quantitatively in space and time. The distribution of airborne pollen from the nine target species occupied distinct temporal windows across the grass pollen

season (Figure 1). Time, as measured by the number of days after the first sample was collected in 2016 and 2017 (25 May 2016 and 5 May 2017, respectively), is a good predictor of airborne grass pollen species composition ($LR_{653,1} = 555.4$, $p = 0.001$; Figure 1). In addition, pollen abundance from different species peaked at different times at each location, with latitude and longitude being good predictors of pollen abundance for each species (latitude, $LR_{652,1} = 68.3$, $p = 0.001$; longitude,
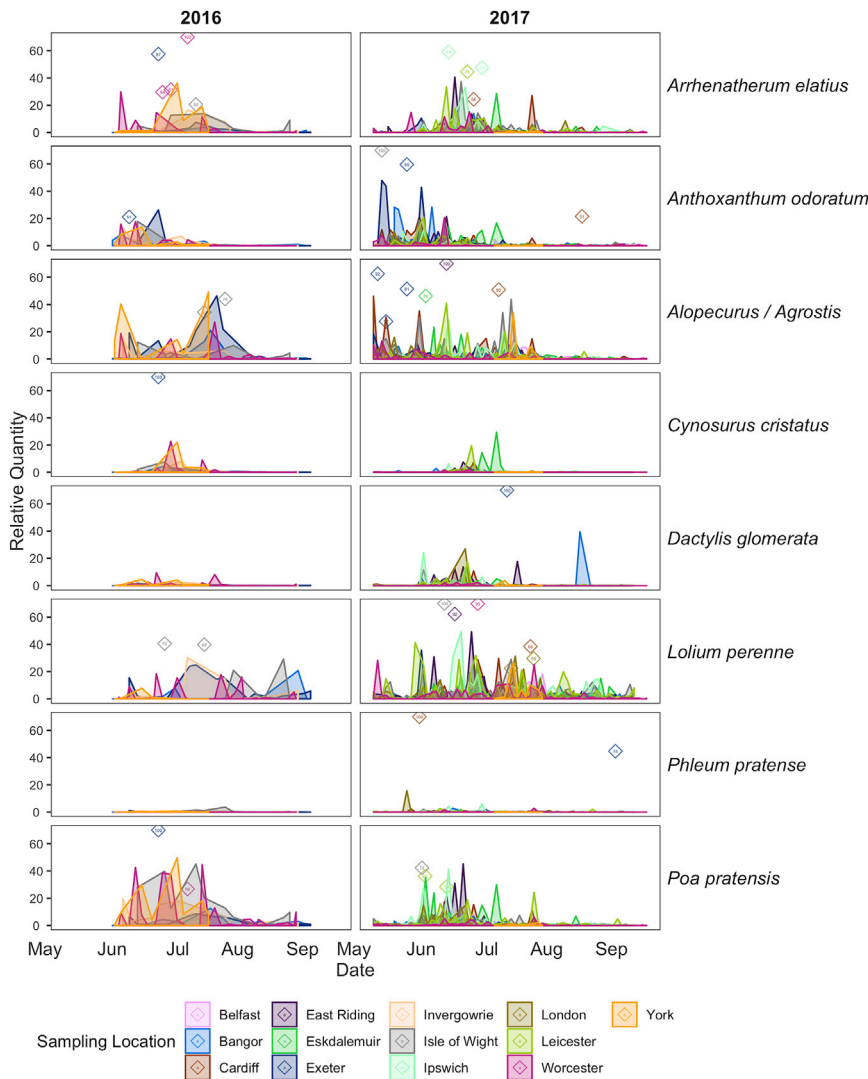
Figure 1. Abundances of airborne pollen from eight grass species during the grass pollen seasons of 2016 and 2017
Abundance of pollen is measured using targeted qPCR and is expressed relative to the maximum abundance for each species. Sampling location is indicated by colored density lines. The y axis is restricted to 60% and the diamonds show pollen abundances higher than 50% with the relative abundance values. See Figure S1 for a map of sampling locations and abundances of grass pollen at each sample location site. See Figure S4 for maximum Poaceae pollen concentrations (maximum measured across one to four consecutive days) at sampling sites.

binary variable, and this probability cannot logically be greater than 100%). However, the confidence intervals suggest that while an association is likely, the effect size is uncertain. In the case of *P. pratense*, the confidence intervals (± 0.84, 2σ) are broader than the maximum effect (0.44), so the reliability of the association itself is uncertain. The model also indicates that high abundances of *Poa pratensis* are associated with decreased probability of occurrence of asthma-related emergency hospital admissions (Figure 2), but the confidence intervals (± 0.38, 2σ) are broader than the maximum effect (-0.24), as with *P. pratense*. Other species (*Arrhenatherum elatius*, *Anthoxanthum odoratum*, *Alopecurus/Agrostis*, *Dactylis glomerata*, and *Lolium perenne*) are not estimated to have species-specific associations. (Effects associated with control variables are presented in Figure 2 and Data S1D. Discussing these in detail is beyond the scope of the present paper.)

$LR_{651,1} = 59$, p = 0.001; Figure S1). Spatial and temporal patterns were not consistent between the grass season in 2016 and 2017 ($LR_{650,1} = 96.1$, p = 0.001; Figure 1).

Associations between hospital admissions and grass species vary depending on the target species
Generalized additive modeling indicates variability in the estimated relative contributions of different grass species (and covariates) on sub-weekly occurrences of asthma-related emergency hospital admissions (STAR Methods) (Figures 2 and S2; Data S1D). In agreement with previous research,[3,4] there are clear associations with overall concentrations of airborne grass pollen. High abundances of the pollen of two grass species, *Cynosurus cristatus* (Crested dog's-tail) and *Phleum pratense* (Timothy), are also associated with increased probability of occurrence of asthma-related emergency hospital admissions. The maximum effect associated with *C. cristatus* is 0.97 (± 0.70, 2σ), which predicts that peak abundances of its pollen are associated with asthma-related emergency admissions being 97% $\pm$ 70% more likely (NB: occurrence of admissions is a

Species-specific associations between monthly prescribing rates and different grass species
We demonstrate that different grass species (and covariates, including overall grass pollen concentrations) have different associations with monthly prescribing rates for drugs used in nasal allergies (BNF 12.2.1) (Figures 3 and S3; Data S1E) and respiratory antihistamines (BNF 3.4.1) (Figures 4 and S3; Data S1F) (STAR Methods). Abundant grass pollen belonging to the genera *Alopecurus* or *Agrostis spp.*, and *Cynosurus cristatus* (crested dog's tail), are associated with higher prescribing rates of drugs used for nasal allergy (Figure 3) and respiratory antihistamines (Figure 4). However, these effects require careful interpretation. The maximum predicted increase in drugs used in nasal allergy prescribing associated with *Alopecurus/Agrostis* pollen is 0.032 (± 0.025, 2σ) items per day per 1,000 population, 1.37% of the total range of nasal allergy drug prescribing rates (minimum: 0.007,
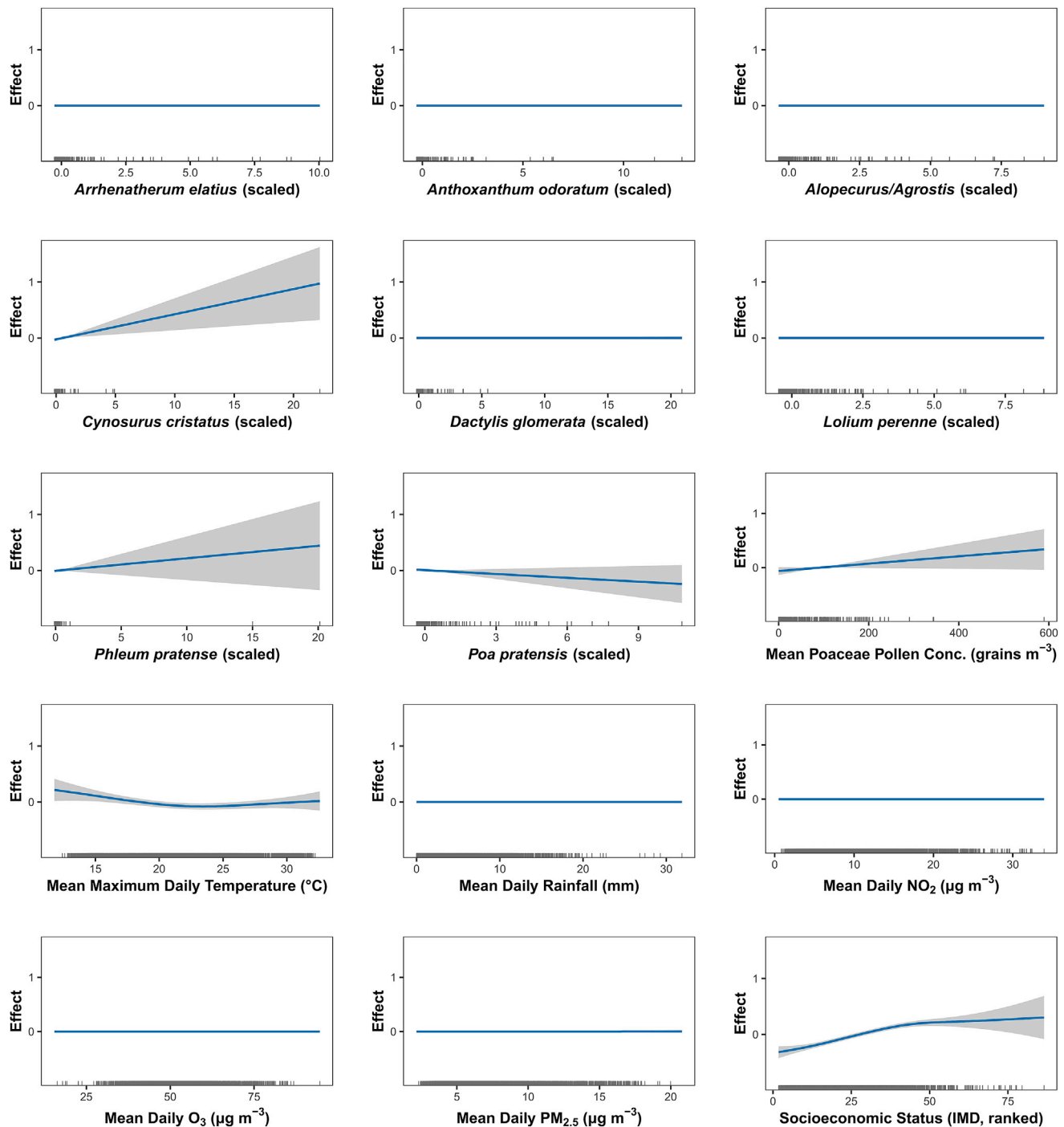
Figure 2. Estimated effects of qPCR target species and covariates on asthma-related emergency hospital admissions

Graphical summary of a generalized additive model (GAM) (Data S1D) showing the estimated effects of major explanatory variables on the occurrence (binary) of emergency asthma-related hospital admissions during the grass pollen season (May to early September) in the UK within 30 km of pollen monitoring stations, with 95% (2σ) confidence intervals (Figure S1). Rug plots indicate x axis values. "Scaled" means the variable has been standardized (zero means and unit variance). (See Figure S2 for unadjusted results and Figure S4 for maximum Poaceae pollen concentrations and other environmental data.)

maximum: 2.343) (Figure 3). For respiratory antihistamines, the maximum predicted increase is 0.047 (±0.050, 2σ) items per day per 1,000 population, 0.98% of the total range of respiratory antihistamine prescribing rates (minimum: 0.013,

maximum: 4.806) (Figure 4). The maximum predicted increase in prescribing drugs used in nasal allergy associated with *C. cristatus* pollen is 0.045 (±0.045, 2σ) items per day per 1,000 population (1.92% of the total range) (Figure 3), and
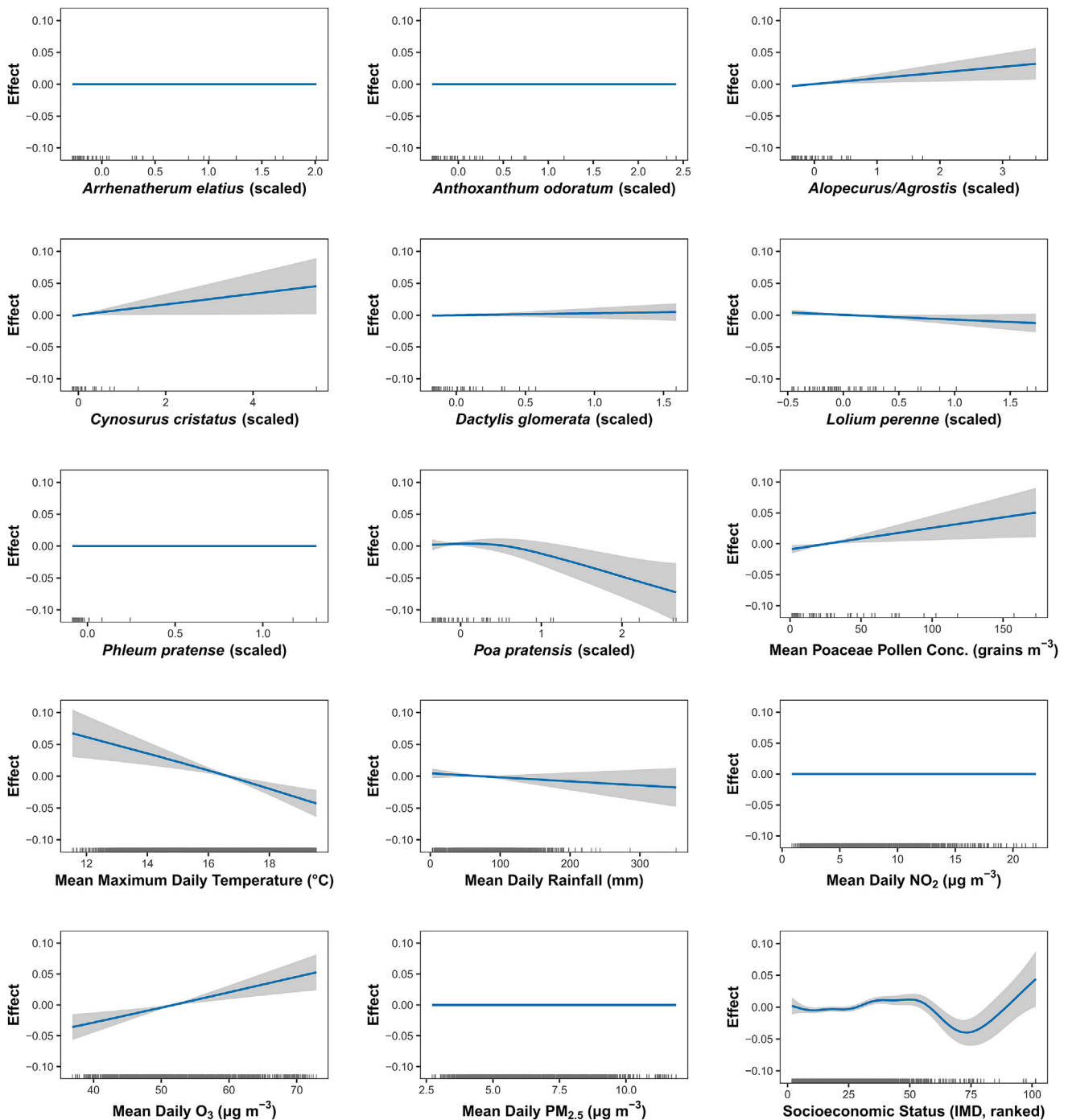
Figure 3. Estimated effects of qPCR target species and covariates on nasal allergy drug prescribing rates

Graphical summary of a generalized additive model (GAM) (Data S1E) showing the estimated effects of major explanatory variables on monthly prescribing rates of drugs for nasal allergy (BNF 12.2.1) per 1,000 population per day during the grass pollen season (May to early September) in the UK within 30 km of pollen monitoring stations, with 95% (2σ) confidence intervals (Figure S1). Rug plots indicate x axis values. "Scaled" means the variable has been standardized (zero means and unit variance). (See Figure S3 for unadjusted results and Figure S4 for maximum Poaceae pollen concentrations and other environmental data.)

for respiratory antihistamines this value is 0.085 (±0.085, 2σ) items per day per 1,000 population (1.77% of the total range) (Figure 4).

Two grass species, *L. perenne* and *Poa pratensis*, are estimated to be associated with lower prescribing rates of both groups of drugs when pollen abundance is elevated, though the association is most notable with *Poa pratensis* (Figures 3 and 4). The maximum predicted change in nasal allergy drug prescribing associated with *Poa pratensis* pollen is −0.073 (±0.046, 2σ) items per day per 1,000 population (Figure 3) (3.12% of the
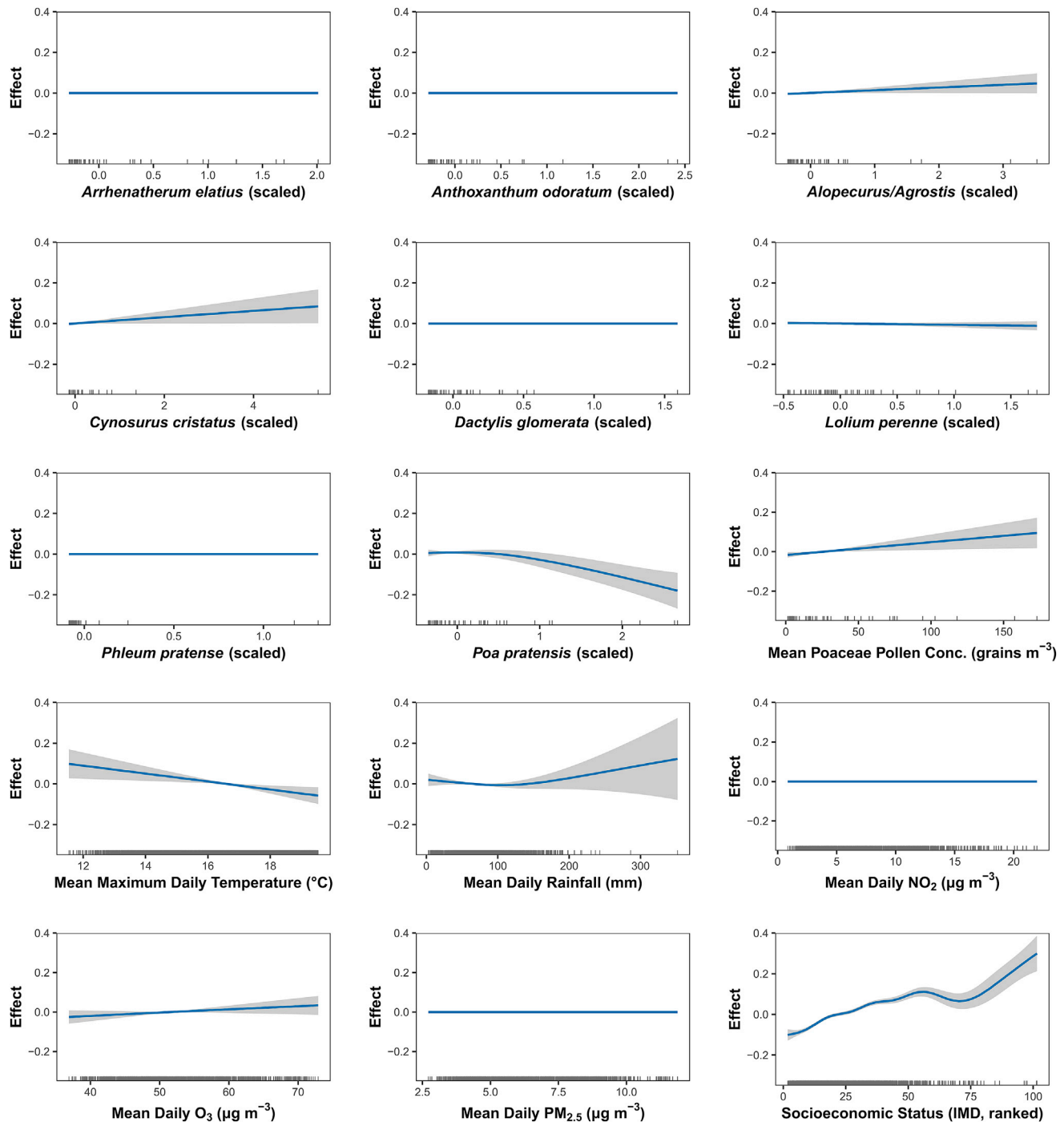
Figure 4. Estimated effects of qPCR target species and covariates on respiratory antihistamine prescribing rates

Graphical summary of a generalized additive model (GAM) (Data S1F) showing the estimated effects of major explanatory variables on monthly prescribing rates of respiratory antihistamines (BNF 3.4.1) per 1,000 population per day during the grass pollen season (May to early September) in the UK within 30 km of pollen monitoring stations, with 95% (2σ) confidence intervals (Figure S1). Rug plots indicate x axis values. "Scaled" means the variable has been standardized (zero means and unit variance). (See Figure S3 for unadjusted results and Figure S4 for maximum Poaceae pollen concentrations and other environmental data.)

total range), and −0.180 (±0.090, 2σ) items per day per 1,000 population for respiratory antihistamines (Figure 4) (3.75% of the total range). The other species (*A. elatius*, *A. odoratum*, *D. glomerata* [BNF 3.4.1], and *Phleum pratense*) are not

estimated to have species-specific effects (Figures 3 and 4). (Effects associated with control variables are presented in Figures 3 and 4 and Data S1. Discussing these in detail is beyond the scope of the present paper.)

Species-level effects of grass pollen on human respiratory health

The grass species studied show different relationships with population respiratory health, demonstrating the potential importance of species-specific grass pollen monitoring. For certain taxa, including *Alopecurus/Agrostis spp.*, *Cynosurus cristatus*, and *Poa pratensis*, the relationships are broadly consistent between health outcome measures (sub-weekly hospital admissions and monthly prescribing rates), and models are also consistent with studies that demonstrate associations between overall grass pollen and population respiratory health response.[3,4] This consistency suggests that predicted relationships are plausible and may reflect reality, but the overall picture warrants discussion. There are a number of reasons why certain taxa may be more (e.g., *Alopecurus/Agrostis spp.*, *C. cristatus*) or less (e.g., *Poa pratensis*) harmful than others (though associations with overall grass pollen concentrations are also consistent between models). Different quantities and types of allergenic proteins (not measured here) have been noted in the pollen of different grass species,[16] and greater concentrations of more potent mixtures of allergenic proteins could result in greater respiratory health impacts associated with certain species, and the reverse for less harmful species. Sensitization is also known to vary between different allergenic proteins in grass pollen.[17] However, while we demonstrate the potential for species-specific effects through associations with health outcomes over two pollen seasons, individual species should not currently be identified as either most or least harmful beyond the specific spatiotemporal context of these models. There may also be undetected taxon-specific effects associated with grass species that were not studied here (though these would be included in overall grass pollen concentrations). Future research would benefit from studying a greater number of species. In addition, while ozone, nitrogen dioxide, and particulate matter ($PM_{2.5}$) have been controlled for in the present study (STAR Methods; Figures 2, 3, and 4; Data S1), other airborne pollutants not included may represent additional, complex sources of uncertainty.[18] For example, certain urban pollutants (e.g., sulfur dioxide, SO_2) may have direct respiratory health impacts,[19] while also decreasing pollen production[20] or pollen allergen production[21] in some plants. Further research is required to understand how intra- and inter-species-level variation in allergens relate to the relative allergenicity and respiratory health impacts of grasses and potential confounding effects of air pollution.

Patient and doctor behavior, particularly with regard to prescribing, is an important and complex variable that must also be considered. There is evidence that GP (general practitioner) consultations for allergic rhinitis are associated with environmental factors[22] and tend to be sought in response to the occurrence of symptoms,[23,24] though some patients seek prescription drugs for respiratory allergies pre-emptively of the grass pollen season. While GP practices, sub-national (e.g., clinical commissioning groups), and national (e.g., NHS Scotland) administrative structures have been included as variables (e.g., "Country") in all GAMs to account for variation in prescribing behavior, it remains a source of uncertainty. Similarly, differences between UK countries in the costs of prescriptions to patients may influence patient and doctor behavior.

Spatiotemporal factors underpin the distribution and taxonomic composition of airborne grass pollen

Understanding the distribution, composition, and abundance of airborne grass pollen is key for understanding associations with human health. We found that variation in species composition is explained by the time of year and location of pollen sampling, which takes into account differences in local meteorological effects and climatic factors associated with plant flowering times.[25] This result is supported by a previous investigation that used eDNA metabarcoding to unravel the community composition of airborne pollen in 2016.[26] Microscopy-based measures of airborne grass pollen abundance, derived from traditional Hirst-type 7-day volumetric samplers, displayed peaks of "high" or "very high" concentrations (>50 grains m^{-3}[4,27]) within a 4-week window between late May and June, and the latter was an especially important month, showing some of the highest airborne grass pollen abundances across the UK (Figure S4).

Some species, such as *C. cristatus*, exhibit singular annual peaks in qPCR-measured abundance in June while other species, such as *L. perenne*, display multiple peaks from May and June onward and some peaks occur as late as July (Figure 1). Multiple peaks may be the result of multiple flowering events over the summer in a single species, such as *L. perenne*, which is the dominant forage crop species, and bred to mature at different times throughout the year.[28] Given that health outcomes are not equal throughout the grass flowering season, it is reasonable to deduct that *Lolium* cultivars do not convey an acute contribution to environmental-derived allergies in the UK. It should be noted that many commercial varieties of *L. perenne* are the result of hybridization with closely related species and molecular-based approaches do not discriminate these species reliably.[26] For example, broad-leaved fescues (currently named *Festuca* or *Schedonorus arundinacea/pratensis*) are genetically very close to and hybridize with *Lolium* spp. and the two groups cannot be genetically distinguished via ITS2/*rbcL* barcode markers, in contrast to narrow-leaved fescues (*Festuca ovina/rubra*), which are clearly distinct.[29] Since *L. perenne* is one of the most abundant grasses in the UK,[26] measuring the abundance of pollen from this species is essential for exploring relationships between grass pollen and human health. The occurrence of multiple peaks of pollen from some qPCR probes may be partly explained by the amplification of non-target species during the qPCR reactions (STAR Methods; Data S1B). For example, *in silico* analysis demonstrates that grass species belonging to the genera *Alopecurus* and *Agrostis* could not be genetically distinguished by the qPCR probes, but they can be separated by annual flowering times (Data S1B). *Agrostis stolonifera* and *Alopecurus pratensis* are both common and widespread grass species in the UK. However, *Agrostis stolonifera* has a later annual flowering time (observed flowering time in June to late August) compared with *Alopecurus pratensis* (one of the earliest observed flowering times in April–May),[30] likely explaining the two peaks of airborne grass pollen in 2016 and 2017 measured in this study using the *Alopecurus/Agrostis spp.* qPCR probe (Figure 1).

We demonstrate the enormous potential of aerial eDNA biomonitoring for studying associations between species-specific pollen abundances and respiratory health responses. However,

while we have captured a broad range of abundances for each species, we are unlikely to have captured the potentially finer-scaled distributions of these values, which would require more temporally resolved sampling (e.g., daily resolution) over a longer period. Furthermore, pollen collection, over multiple years, is required in order to make accurate predictions about the spatio-temporal composition of airborne pollen. The pollen data are also assumed to be representative within a 30 km radius, which is a reasonable "rule of thumb,"[31–33] but a more spatially resolved sampling would be preferable in future research.

Conclusions and future prospects

The research presented here represents the first of its kind to use a nationwide, ecological approach to biomonitoring for the purpose of exploring relationships between airborne pollen and human health. We demonstrate that different species of airborne grass pollen are associated with differing respiratory health responses, as well as overall concentrations of grass pollen. The data suggest that certain species may have disproportionate impacts, relative to overall grass pollen concentrations, potentially as a result of species-specific allergen mixtures, while other species may have lower impacts. We show that grass pollen assemblages are quantitatively, spatiotemporally structured during each pollen season, and health impacts may be further moderated on this basis. We envisage the development of a global network of autonomous aerial samplers, able to discriminate and quantify airborne pollen, allowing sensitive biomonitoring of important aeroallergens at high spatial and temporal resolutions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead Contact
  - ○ Materials Availability
  - ○ Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Airborne pollen sampling
  - ○ DNA extraction and quantitative PCR (qPCR)
  - ○ Public health data and study population
  - ○ Additional variables
- QUANTIFICATION AND STATISTICAL ANALYSIS

AUTHOR CONTRIBUTIONS

S.C., N.J.O., C.A.S., G.W.G., R.N.M., Y.C., B.W., A.W., N.d.V., F.M.R., and G.L.B. conceived and designed the study; B.A.-G., J.A., J.B., G.L.B., A.E., S.H., K.A.L., R.N., C.H.P., G.M.P., R.N., S.P., H.M.R., K.S., T.E.L.S., N.S., J.T., and J.Z.-C. collected samples and counted pollen; G.L.B. and C.P. performed DNA extractions and G.L.B. performed qPCR lab work, supported by S.C.; F.M.R., T.E., R.N.M., N.d.V., L.J., G.L.B., and S.C. contributed methods; F.M.R. and G.L.B. analyzed the data; and F.M.R., G.L.B., N.J.O., B.W., and S.C. produced the first draft of the manuscript. All authors contributed to the final submitted manuscript.

REFERENCES

1. Burbach, G.J., Heinzerling, L.M., Edenharter, G., Bachert, C., Bindslev-Jensen, C., Bonini, S., Bousquet, J., Bousquet-Rouanet, L., Bousquet, P.J., Bresciani, M., et al. (2009). GA(2)LEN skin test study II: clinical relevance of inhalant allergen sensitizations in Europe. Allergy 64, 1507–1515.

2. Bousquet, J., Schünemann, H.J., Samolinski, B., Demoly, P., Baena-Cagnani, C.E., Bachert, C., Bonini, S., Boulet, L.P., Bousquet, P.J., Brożek, J.L., et al.; World Health Organization Collaborating Center for Asthma and Rhinitis (2012). Allergic Rhinitis and its Impact on Asthma (ARIA): achievements in 10 years and future needs. J. Allergy Clin. Immunol. 130, 1049–1062.

3. Erbas, B., Chang, J.H., Dharmage, S., Ong, E.K., Hyndman, R., Newbigin, E., and Abramson, M. (2007). Do levels of airborne grass pollen influence asthma hospital admissions? Clin. Exp. Allergy 37, 1641–1647.

4. Osborne, N.J., Alcock, I., Wheeler, B.W., Hajat, S., Sarran, C., Clewlow, Y., McInnes, R.N., Hemming, D., White, M., Vardoulakis, S., and Fleming, L.E. (2017). Pollen exposure and hospitalization due to asthma exacerbations: daily time series in a European city. Int. J. Biometeorol. 61, 1837–1848.

5. Cecchi, L. (2012). Introduction. In Allergenic Pollen: A Review of the Production, Release, Distribution and Health Impacts, M. Sofiev, and K.-C. Bergmann, eds. (Springer), pp. 1–7.

6. Bousquet, J., Dahl, R., and Khaltaev, N. (2007). Global alliance against chronic respiratory diseases. Allergy 62, 216–223.

7. Marks, G., Pearce, N., Strachan, D., and Asher, I. (2014). Global burden of disease due to asthma. The Global Asthma Report 2014 (The Global Asthma Network), pp. 16–21.

8. Cruz, A.A., Popov, T., Pawankar, R., Annesi-Maesano, I., Fokkens, W., Kemp, J., Ohta, K., Price, D., and Bousquet, J.; ARIA Initiative Scientific Committee (2007). Common characteristics of upper and lower airways in rhinitis and asthma: ARIA update, in collaboration with GA(^{2})LEN. Allergy 62 (*Suppl 84*), 1–41.

9. NHS (National Health Service) (2017). Specialised respiratory services (adult) – severe asthma. https://www.england.nhs.uk/publication/specialised-respiratory-services-adult-severe-asthma/.

10. Menzel, A. (2019). The allergen riddle. Nat. Ecol. Evol. *3*, 716–717.

11. Mander, L., Li, M., Mio, W., Fowlkes, C.C., and Punyasena, S.W. (2013). Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. Proc. Biol. Sci. *280*, 20131905.

12. Müller-Germann, I., Vogel, B., Vogel, H., Pauling, A., Fröhlich-Nowoisky, J., Pöschl, U., and Després, V.R. (2015). Quantitative DNA analyses for airborne birch pollen. PLoS ONE *10*, e0140949.

13. Kraaijeveld, K., de Weger, L.A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P.S., and den Dunnen, J.T. (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. Mol. Ecol. Resour. *15*, 8–16.

14. Campbell, B.C., Al Kouba, J., Timbrell, V., Noor, M.J., Massel, K., Gilding, E.K., Angel, N., Kemish, B., Hugenholtz, P., Godwin, I.D., and Davies, J.M. (2020). Tracking seasonal changes in diversity of pollen allergen exposure: targeted metabarcoding of a subtropical aerobiome. Sci. Total Environ. *747*, 141189.

15. Banchi, E., Ametrano, C.G., Tordoni, E., Stanković, D., Ongaro, S., Tretiach, M., Pallavicini, A., and Muggia, L.; ARPA Working Group (2020). Environmental DNA assessment of airborne plant and fungal seasonal diversity. Sci. Total Environ. *738*, 140249.

16. Chabre, H., Gouyon, B., Huet, A., Baron-Bodo, V., Nony, E., Hrabina, M., Fenaille, F., Lautrette, A., Bonvalet, M., Maillère, B., et al. (2010). Molecular variability of group 1 and 5 grass pollen allergens between Pooideae species: implications for immunotherapy. Clin. Exp. Allergy *40*, 505–519.

17. Hrabina, M., Peltre, G., Van Ree, R., and Moingeon, P. (2008). Grass pollen allergens. Clin. Exp. Allergy Rev. *8*, 7–11.

18. Frank, U., and Ernst, D. (2016). Effects of NO_2 and ozone on pollen allergenicity. Front. Plant Sci. *7*, 91.

19. Guarnieri, M., and Balmes, J.R. (2014). Outdoor air pollution and asthma. Lancet *383*, 1581–1592.

20. Bosac, C., Black, V.J., Black, C.R., Roberts, J.A., and Lockwood, F. (1993). Impact of O_3 and SO_2 on reproductive development in oilseed rape (Brassica napus L.). I. Pollen germination and pollen tube growth. New Phytol. *124*, 439–446.

21. Rogerieux, F., Godfrin, D., Sénéchal, H., Motta, A.C., Marlière, M., Peltre, G., and Lacroix, G. (2007). Modifications of Phleum pratense grass pollen allergens following artificial exposure to gaseous air pollutants ($O_{(3)}$, $NO_{(2)}$, $SO_{(2)}$). Int. Arch. Allergy Immunol. *143*, 127–134.

22. Todkill, D., de Jesus Colon Gonzalez, F., Morbey, R., Charlett, A., Hajat, S., Kovats, S., Osborne, N.J., McInnes, R., Vardoulakis, S., Exley, K., et al. (2020). Environmental factors associated with general practitioner consultations for allergic rhinitis in London, England: a retrospective time series analysis. BMJ Open *10*, e036724.

23. Todkill, D., Loveridge, P., Elliot, A.J., Morbey, R., Lusignan, S., Edeghere, O., and Smith, G. (2017). Socioeconomic and geographical variation in general practitioner consultations for allergic rhinitis in England, 2003-2014: an observational study. BMJ Open *7*, e017038.

24. Green, H.K., Edeghere, O., Elliot, A., Cox, I., Mckendry, R., and Smith, G. (2018). What value can Google search data add to existing syndromic surveillance systems? ISDS Annu. Conf. Proc *10*, 2579.

25. Emberlin, J., Jones, S., Bailey, J., Caulton, E., Corden, J., Dubbels, S., Evans, J., McDonagh, N., Millington, W., Mullins, J., et al. (1994). Variation in the start of the grass pollen season at selected sites in the United Kingdom 1987-1992. Grana *33*, 94–99.

26. Brennan, G.L., Potter, C., de Vere, N., Griffith, G.W., Skjøth, C.A., Osborne, N.J., Wheeler, B.W., McInnes, R.N., Clewlow, Y., Barber, A., et al.; PollerGEN Consortium (2019). Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. Nat. Ecol. Evol. *3*, 750–754.

27. Met Office (2020). What is the pollen count? https://www.metoffice.gov.uk/weather/warnings-and-advice/seasonal-advice/health-wellbeing/pollen/what-is-the-pollen-count.

28. Emecz, T.I. (1962). The effect of meteorological conditions on anthesis in agricultural grasses. Ann. Bot. *26*, 159–172.

29. Guan, X., Yuyama, N., Stewart, A., Ding, C., Xu, N., Kiyoshi, T., and Cai, H. (2017). Genetic diversity and structure of *Lolium* species surveyed on nuclear simple sequence repeat and cytoplasmic markers. Front. Plant Sci. *8*, 584.

30. Cope, C., and Gray, A. (2009). Grasses of the British Isles (Botanical Society of Britain and Ireland).

31. Pashley, C.H., Fairs, A., Edwards, R.E., Bailey, J.P., Corden, J.M., and Wardlaw, A.J. (2009). Reproducibility between counts of airborne allergenic pollen from two cities in the East Midlands, UK. Aerobiologia (Bologna) *25*, 249–263.

32. Skjøth, C.A., Smith, M., Šikoparija, B., Stach, A., Myszkowska, D., Kasprzyk, I., Radišić, P., Stjepanović, B., Hrga, I., Apatini, D., et al. (2010). A method for producing airborne pollen source inventories: an example of *Ambrosia* (ragweed) on the Pannonian Plain. Agric. For. Meteorol. *150*, 1203–1210.

33. Skjøth, C.A., Ørby, P.V., Becker, T., Geels, C., Schlünssen, V., Sigsgaard, T., Bønløkke, J.H., Sommer, J., Søgaard, P., and Hertel, O. (2013). Identifying urban sources as cause of elevated grass pollen concentrations using GIS and remote sensing. Biogeosciences *10*, 541–554.

34. European Centre for Environment & Human Health (2019). MEDMI – connecting health and environment data. https://www.data-mashup.org.uk/.

35. NHS (National Health Service) Digital (2018). GP and GP practice related data. https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data.

36. NHS BSA (National Health Service Business Services Authority) (2018). Information Services Portal (ISP). https://www.nhsbsa.nhs.uk/information-services-portal-isp.

37. NHS (National Health Service) Wales Shared Services Partnership (2018). General practice prescribing data extract. http://www.primarycareservices.wales.nhs.uk/general-practice-prescribing-data-extrac.

38. NHS (National Health Service) National Services Scotland (2018). Prescriptions in the community. https://www.opendata.nhs.scot/dataset/prescriptions-in-the-community.

39. Open Data Northern Ireland (2018). GP prescribing data. https://www.opendatani.gov.uk/dataset/gp-prescribing-data.

40. Met Office (2018). HadUK-Grid gridded and regional average climate observations for the UK. http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb.

41. Copernicus Atmosphere Modelling Service (2018). European air quality. http://www.regional.atmosphere.copernicus.eu/?category=data_access.

42. Abel, G.A., Barclay, M.E., and Payne, R.A. (2016). Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. BMJ Open *6*, e012750.

43. O'Brien, O. (2016). Population density and urban/rural classification. https://data.cdrc.ac.uk/dataset/population-density-and-urban-rural-classification.

44. ONS (Office for National Statistics) (2017). Rural urban classification (2011) of lower layer super output areas in England and Wales. http://geoportal.statistics.gov.uk/datasets/rural-urban-classification-2011-of-lower-layer-super-output-areas-in-england-and-wales.

45. de Vere, N., Rich, T.C.G., Ford, C.R., Trinder, S.A., Long, C., Moore, C.W., Satterthwaite, D., Davies, H., Allainguillaume, J., Ronca, S., et al. (2012). DNA barcoding the native flowering plants and conifers of Wales. PLoS ONE *7*, e37945.

46. Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C., and Bik, H.M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. Methods Ecol. Evol. *7*, 1008–1018.

47. Tobías, A., Galán, I., and Banegas, J.R. (2004). Non-linear short-term effects of airborne pollen levels with allergenic capacity on asthma emergency room admissions in Madrid, Spain. Clin. Exp. Allergy *34*, 871–878.

48. Wheeler, B.W., Gunnell, D., Metcalfe, C., Stephens, P., and Martin, R.M. (2008). The population impact on incidence of suicide and non-fatal self harm of regulatory action against the use of selective serotonin reuptake inhibitors in under 18s in the United Kingdom: ecological study. BMJ *336*, 542–545.

49. Rowlingson, B., Lawson, E., Taylor, B., and Diggle, P.J. (2013). Mapping English GP prescribing data: a tool for monitoring health-service inequalities. BMJ Open *3*, e001363.

50. Wheeler, B.W., and Ben-Shlomo, Y. (2005). Environmental equity, air quality, socioeconomic status, and respiratory health: a linkage analysis of routine data from the Health Survey for England. J. Epidemiol. Community Health *59*, 948–954.

51. Milojevic, A., Niedzwiedz, C.L., Pearce, J., Milner, J., MacKenzie, I.A., Doherty, R.M., and Wilkinson, P. (2017). Socioeconomic and urban-rural differentials in exposure to air pollution and mortality burden in England. Environ. Health *16*, 104.

52. R Core Team (2019). R: A language and environment for statistical computing.

53. Wang, Y., Naumann, U., Wright, S.T., and Warton, D.I. (2012). Mvabund – an R package for model-based analysis of multivariate abundance data. Methods Ecol. Evol. *3*, 471–474.

54. Birks, H.J.B. (2012). Overview of numerical methods in palaeolimnology. In Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques, H.J.B. Birks, A.F. Lotter, S. Juggins, and J.P. Smol, eds. (Springer), pp. 19–92.

55. Wood, S.N. (2017). Generalized Additive Models: An Introduction with R, 2nd Edition (CRC Press/Taylor & Francis Group).

56. Wood, S.N. (2019). mgcv (R Project).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
qPCR data	NERC EIDC	https://doi.org/10.5285/28208be4-0163-45e6-912c-2db205126925
Oligonucleotides		
Primers and probes manufactured by Primer Design	Data S1A	N/A

RESOURCE AVAILABILITY

Lead Contact
The lead contact for resource availability is Georgina Brennan (g.l.b.doonan@gmail.com).

Materials Availability
This study did not generate new unique reagents.

Data and Code Availability
Original qPCR data have been deposited to NERC EIDC: https://doi.org/10.5285/28208be4-0163-45e6-912c-2db205126925. Standard pollen monitoring 'count' data were sourced from the MEDMI database,[34] with the exception of data from Bangor which were produced as part of the present study and are available on request. Prescribing datasets are publicly available,[35–39] as are weather,[40] air pollution,[41] deprivation (IMD)[42,43] and rural-urban category[43,44] data. Hospital episode statistics (HES) datasets are sensitive, individual-level health data, which are subject to strict privacy regulations and are not publicly available. The study did not generate any unique code.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study uses abundance data (derived from qPCR analysis) of airborne pollen including *Anthoxanthum odoratum* (sweet vernal-grass), *Arrhenatherum elatius* (false oat-grass), *Cynosurus cristatus* (crested dog's-tail), *Dactylis glomerata* (cock's-foot), *Lolium perenne* (perennial ryegrass), *Phleum pratense* (Timothy), *Poa pratensis* (smooth meadow-grass), grass species within the genera *Alopecurus/Agrostis*, and one probe that was found to be degenerate and unable to discriminate grass species ('Poaceae indet.'). We used qPCR to track the seasonal progression of airborne grass pollen, in time and space. To do this we collected aerial samples from thirteen sites across the UK during the pollen seasons (May to September) of 2016 and 2017. Two proxies of respiratory health responses were used in this study: prescribing rates for drugs used in treating respiratory allergies, and records of asthma-related emergency hospital admissions. Full details are provided in the following sections.

METHOD DETAILS

Airborne pollen sampling
Airborne pollen was collected from 13 sites across the UK (Figure S1). The sites are permanent pollen monitoring stations, apart from Bangor, which supply pollen count information for the UK forecast. Airborne pollen samples were collected using Burkard Automatic Multi-Vial Cyclone Samplers (V2; Burkard Manufacturing) which are similar to the Hirst design (Burkard Manufacturing), a.k.a. the seven-day volumetric trap. However, instead of a collection onto an adhesive-coated tape for microscopy, the aerial particles are collected into 1.5 mL sterile microcentrifuge and DNA from the entire contents of the tubes are extracted for downstream molecular analysis. In 2016, pollen was collected at 6 sites, then in 2017 a further 7 sites were added (n.b. samples for molecular analysis were not collected from Invergowrie in 2017). Daily samples were taken during the grass pollen seasons (samples were collected between 25th May to 28th August 2016 and 5th May to 10th September 2017) (Brennan et al.[26]; STAR methods).

DNA extraction and quantitative PCR (qPCR)
Pollen collected in 2016 was extracted as described in Brennan et al.[26] Daily airborne pollen samples were pooled over three consecutive days and on one occasion four days. Due to the large sample sizes, slight modifications to the DNA extraction protocol were made for the extraction of pollen from samples collected in 2017 that were extracted using DNeasy 96 Plant Kits (QIAGEN, Valencia, CA, USA). Of 1,400 daily aerial samples, 1,210 were selected for downstream molecular analysis (samples were excluded if pollen

could not be reliably extracted due to large volumes of rainwater in collection tubes). Daily samples were pooled into consecutive 48 h samples at the binding stage of the DNeasy 96 Plant Kit and eluted into 120 μl of elution buffer, yielding 605 pools of DNA for downstream analysis.

Quantitative polymerase chain reaction (qPCR) was performed using a QuantStudio 6 Flex Real-Time qPCR machine (Thermo-Fisher Scientific). Species-specific primers were designed to target the ITS2 region of abundant grass species in the UK including *Anthoxanthum odoratum* (sweet vernal-grass), *Arrhenatherum elatius* (false oat-grass), *Cynosurus cristatus* (crested dog's-tail), *Dactylis glomerata* (cock's-foot), *Lolium perenne* (perennial ryegrass), *Phleum pratense* (Timothy), *Poa pratensis* (smooth meadow-grass), grass species within the genera *Alopecurus/Agrostis*, and one primer set that was found to be degenerate and unable to discriminate grass species (Poaceae indet.) (Data S1A). The ITS2 region was selected as it is multiple copy and all UK native species have DNA reference barcodes for both the ITS2 and *rbcL* markers.[45] Preliminary experiments using single copy qPCR primers/probes and comparisons with metabarcoding data generated from the same DNA samples confirmed that single copy primers were unreliable for detecting target species at low concentrations (low concentration of DNA is common in eDNA samples). For quantification, dilution series of a copy number DNA control carrying the target sequence were used with each species specific primer / probe. Negative controls were run under the same conditions, but nuclease-free water replaced DNA in these wells. Each 10 μL qPCR reaction contained 1 × PrecisionPLUS qPCR Master Mix, with ROX at a lower level (PPLUS-LR, Primer Design, UK), 6 μmol L^{-1} species specific probe and forward and reverse primer, 0.5 μL of DNA template and 4 μl of nuclease-free water. The thermocycling program began with an initial 95°C step for 2 min followed by 50 cycles of 10 s at 95°C and 1 min at 60°C, as per the manufacturer's instructions.

Values produced for each species using qPCR methods are not comparable on an inter-species basis in terms of their relation to biomass (i.e., pollen grains or individual flowering plants).[46] These data were therefore standardized (zero means and unit variance, i.e., 'z-scores') prior to statistical analyses involving public health data (values reflect the number of standard deviations from the mean and this is referred to as 'scaled' in axis labels). This facilitates analyses on the basis of the direction and relative magnitude of changes in taxon abundances. For abundance analysis of qPCR data, where negative values cannot be used, the quantity of each target species was normalized relative to the maximum value for each species, across both 2016 and 2017. See Data S1C for summary statistics.

qPCR primers/probes were designed to be species-specific. However, there are instances where additional grass species may be detected (STAR methods; Data S1B), yet, this does not prevent us from addressing our focal questions. The risks of false detection can be assessed *in silico* by comparing the sequence of the primers and probes with the target species and closely related species found in the UK alongside consideration of the phenology (i.e., flowering time) and rarity of grasses with similar primer and probe target sequences (Data S1B). For instance, the *Cynosurus cristatus* probe may also detect the closely related *Catapodium rigidum* (fern grass). However, the abundance/distribution of *C. rigidum* is much more limited than *C. cristatus*.[30] Furthermore, its pollen was only detected in two of the samples analyzed in Brennan et al.,[26] and at much lower levels than *C. cristatus*. Based on this evidence, it is very unlikely that we are detecting *C. rigidum* in our environmental samples when using the primer/probe designed to detect *C. cristatus* (Data S1B).

Public health data and study population
Two proxies of respiratory health responses were used in this study: prescribing rates for drugs used in treating respiratory allergies, and records of asthma-related emergency hospital admissions.

Prescribing rates for two categories of drugs were used in the analyses: respiratory antihistamines (British National Formulary (BNF 3.4.1) (e.g., Cetirizine Hydrochloride) and drugs used in nasal allergy (BNF 12.2.1) (e.g., Beclometasone Dipropionate). There is no overlap between these categories of drugs. Prescribing rates were calculated as items (the number of times a product appears on a prescription) prescribed per 1000 population per GP practice per month, using GP list sizes as a proxy for population (data were obtained from open sources[35–39]). These rates were divided by the number of days in the month (to account for varying numbers of days each month), providing monthly averages of items prescribed per day per 1000 population. Equivalent data are not available at finer temporal resolutions at population scale. However, the data presented here are at population scale and are highly spatially resolved (per GP practice), allowing analysis of less severe respiratory health responses at population scale, complimentary to parallel analyses involving more temporally resolved hospital admissions data. See Data S1C for summary statistics.

Data for emergency asthma-related admissions were obtained from hospital episode statistics (HES) datasets, from NHS Digital and NHS Wales Informatics Service. These data were filtered to derive daily counts of admissions per GP practice, age category (0-5, 6-18, 19-64, 65+) and sex, with primary diagnosis assigned to ICD (International Classification of Disease) code J45 (asthma) or J46 (acute severe asthma) (England and Wales only). See Data S1C for summary statistics.

The two to five day lag between high pollen concentration exposure and asthma exacerbation[4,47] must be accounted for. Doing so is not straightforward given the constraints imposed by the uneven temporal structure of the pollen dataset (airborne pollen was collected over 24 h periods, and these samples were combined into one to four day 'pools' for eDNA analysis), and standard time series analytical approaches cannot be used. A blanket three-day lag was therefore applied to the admissions data prior to analyses, which was achieved by moving admission dates backward three days to coincide more closely with relevant exposures. Admissions counts were then grouped according to the temporal structure of the pollen dataset (one to four-day pooled airborne pollen samples)

and treated as a binary variable indicating either the presence or absence of emergency asthma admissions within each set of one to four days. This results in a complex temporal structure, with resolution ranging from daily to four-daily. We refer to this with the term 'sub-weekly' (i.e., finer resolution than weekly).

GP practices within 30 km of pollen monitoring stations (Figure 1) were used as the primary geographic units. Airborne pollen concentrations measured at meteorological stations are considered to be reasonably representative of concentrations within 30 km.[31–33] Analyses using prescribing data included pollen measurements from all 13 monitoring stations, while for analyses involving hospital admissions data only the 10 stations in England and Wales were used (admissions data were only available for England and Wales).

Prescribing and hospital admissions datasets derive from National Health Service (NHS) primary and secondary care activity, respectively. Since the NHS accounts for a large proportion of clinical activity in the UK, it is reasonable to assume that the private-sector healthcare excluded here is a negligible omission, and NHS data resources are frequently used for population-scale studies.[4,48,49]

Additional variables

Additional variables were included in the respiratory health models to adjust for potential confounding effects of the associations between pollen abundances, hospital admissions and prescribing rates. We included variables relating to social, demographic and behavioral factors at unit (age category, sex) and population levels (deprivation indices, rural-urban classification, day of the week), as well as those relating to air quality and the physical environment (total airborne grass pollen concentrations, temperature, precipitation, nitrogen dioxide, ozone, particulate matter). Age category, sex and day of the week are only included in analyses involving hospital admissions data. Where appropriate, data have been averaged according to the temporal structures of each public health dataset (e.g., for variables with sub-monthly temporal resolutions, location-specific monthly mean values were calculated prior to joining with monthly prescribing datasets). See Data S1C for summary statistics.

Daily concentrations of airborne grass pollen averaged according to the temporal structures of each type of data (monthly for prescribing data, and sub-weekly for admissions data), were included in the models alongside species-specific qPCR data. Daily air concentrations of grass pollen (derived from traditional sampling using Hirst-type seven-day volumetric samplers) were sourced from the MEDMI database,[34] from data provided by the Met Office (the UK's meteorological service), with the exception of data from Bangor which were produced as part of the present study.

Monthly values of mean air temperature (°C) and total precipitation (mm) were used with prescribing data, and daily values of maximum air temperature (°C) and total precipitation (mm) were used with admissions data. Openly available, interpolated observational weather datasets (1 km grids)[40] were used in these analyses. The daily data were grouped to correspond with multi-day pollen sample 'pools' and averaged across these days. Weather data were linked spatially to health data according to the nearest cell to each GP practice (usually, the cell in which GP practice postcode centroids are located).

Daily mean values (averaged for monthly health data) of nitrogen dioxide (NO_2), ozone (O_3) and particulate matter ($PM_{2.5}$) were derived from ensemble model datasets on a 0.1° grid, based on interpolated observations[41] and associations between pollen and health outcomes adjusted for these pollutants. Daily values used with admissions data were grouped to correspond with multi-day pollen sample 'pools' and averaged across these days. Spatially, these were linked to health data according to the nearest cell coordinates. Points derived from NetCDF coordinates represent southwestern corners of grid cells, rather than centroids, but given the grid's fine scale (0.1°, which is ~6-11 km for the UK) this is a reasonable pragmatic compromise for a non-critical dataset, which was necessary to efficiently link datasets with different underlying architectures.

Social environmental factors which influence health, such as rural-urban status and socioeconomic status,[50,51] were accounted for. For administrative reasons, urban-rural classifications and indices of multiple deprivation (IMD) are determined separately in each of the countries within the UK. In order to conduct analyses involving multiple UK countries, common measures of these variables must be used. We use openly available IMD rankings (higher scores relate to greater deprivation) for lower layer super output areas (LSOAs) for the whole UK, which have been based primarily on income and employment metrics.[42] For analyses involving prescribing data, we use openly available LSOA-level urban-rural classifications for the whole UK, developed by unifying categories from different administrations.[43] Official, unified LSOA-level rural-urban classifications for England and Wales are openly available,[44] and these are used in analyses involving hospital admissions.

QUANTIFICATION AND STATISTICAL ANALYSIS

Following molecular data collection, pollen qPCR data were analyzed using QuantStudio Software V1.3 (Thermo Fisher Scientific). The number of copies for each target species was quantified using copy number standards provided by Primer Design (standard curves ranged from 2 to 2×10^5 copies per μl for each species). Quantitative PCR runs with PCR efficiencies less than 85% and greater than 115% were not used for further analysis (efficiency of qPCR data used in downstream analysis ranged between 88.5% and 106%). The variability of all molecular data was quantified in R (version 3.6.0) and data points with a large standard deviation between three technical replicates (> 6.95, based on the upper quartile range of the data) were removed. In addition, samples which amplified before 10 cycles and after 38 cycles were removed to reduce the chance of detecting false positive or false negative amplification respectively.

To understand how the grass pollen composition changed quantitatively with space and time, the effect of time (measured as the number of days after the first sampling date for 2016 and 2017), year of sample collection, latitude and longitude of sampling location

were included in a two-tailed generalized linear model (GLM) using the 'manyglm' function in the R[52] package 'mvabund.'[53] The quantity of each grass species (collected using qPCR) was set as the response variable. The effects of time, year, latitude and longitude were included as explanatory variables in the models, specifically assuming a linear association with each covariate and the mean of the response. Overfitting of the models was tested using 'dropterm' in R and based on the lowest Akaike Information Criterion (AIC) score and the Bayesian information criterion (BIC) score. In addition, the appropriateness of the models was checked by visual inspection of the residuals against predicted values from the models. The R code to run these models with 'mvabund'[53] followed a general form, outlined here:

manyglm(mvadbund object = relative abundance of grass taxon ~Time + Latitude + Longitude, family = "negative binomial," metadata)

Generalized additive models (GAMs) are employed here as the primary tool for statistical analyses involving health data. These are highly flexible models, which do not assume linearity in the relationships between the mean of the response and the explanatory variables.[54,55] GAMs are an extension of generalized linear models (GLMs), with the exception that the effects of the explanatory variables on the mean response are defined by 'unknown' smooth functions that are estimated from the data, rather than being assumed to be known (e.g., linear) *a priori*. As such, GAMs may be considered data-driven as opposed to model-driven.[54] These analyses were carried out using the R package 'mgcv.'[56] This particular implementation of GAMs avoids overfitting by objectively penalizing the unknown functions as a compromise between in-sample and out-of-sample predictive accuracy.

The construction of GAMs for each type of response variable (prescribing rates and hospital admissions), was broadly similar. All explanatory variables were included in each, as well as random effects for pollen monitoring site, primary care organization and country (i.e., England, Scotland, Wales and Northern Ireland), and categorical terms to account for variability attributed to month, week of the year (i.e., 1-52), and weekend days (*n.b.* the latter two only apply to daily admissions data). GP practices (n≫2500) were initially treated as individual random effects, but this resulted in highly inefficient models, and so GP practice postcode coordinates (latitude and longitude) were used to treat these as a spatial (smooth) surface instead. This results in more efficient models, and also allows the model to account for spatial autocorrelation between GP practices (practice catchments overlap geographically, and therefore proximity will result in similarity). Prescribing data (for both BNF 3.4.1 and BNF 12.2.1) are 'heavy-tailed' compared to the Normal distribution, and so a scaled t-distribution was utilized. Emergency asthma-related hospital admissions are a rare occurrence, particularly during pollen season months (May to September), and so these data are treated as binary (i.e., occurrence or no occurrence) and a binomial distribution was used (with the covariates influencing the probability of occurrence at the logistic level). The R code to run these models with 'mgcv'[56] followed a general form, outlined here (*n.b.* models based on prescribing datasets used the 'gam()', family = "scat," and did not include 'Days', 'Weekend', Week', 'Sex' or 'AgeCategory'):

bam(HealthOutcome ~s(A.elatius, bs = "ts") + s(A.odoratum, bs = "ts") + s(Alopecurus_Agrostis, bs = "ts") + s(C.cristatus, bs = "ts") + s(D.glomerata, bs = "ts") + s(L.perenne, bs = "ts") + s(P.pratense, bs = "ts") + s(P.pratensis, bs = "ts") + s(PoaceaeIndet, bs = "ts") + s(MaxPoaceaeConc, bs = "ts") + RuralUrban + s(IMD, bs = "ts") + s(Temperature, bs = "ts") + s(Rainfall, bs = "ts") + s(NO2, bs = "ts") + s(O3, bs = "ts") + s(PM2.5, bs = "ts") + s(Site.ID, bs = "re") + s(GP_Lat, GP_Long, bs = "ts," k = 50) + s(CCG, bs = "re") + s(Country, bs = "re") + Days + Weekend + s(Week, bs = "re") + Month + Year + Sex + AgeCategory, discrete = TRUE, nthreads = 8, data = hes, family = "binomial," method = "fREML")