



Crowd-sourced allergic rhinitis symptom data: The influence of environmental and demographic factors

Jeremy D. Silver^a, Kymble Spriggs^{b,c}, Simon Haberle^{d,e}, Constance H. Katelaris^f, Edward J. Newbigin^g, Edwin R. Lampugnani^{g,*}

^a School of Earth Sciences, University of Melbourne, Parkville, Victoria, Australia

^b Melbourne Medical School, University of Melbourne, Parkville, Victoria, Australia

^c Department of Allergy and Immunology, Royal Melbourne Hospital, Parkville, Victoria, Australia

^d School of Culture, History and Language, College of Asia and the Pacific, Australian National University Canberra, Canberra, Australian Capital Territory, Australia

^e ARC Centre of Excellence for Australian Biodiversity and Heritage, Australian National University, Canberra, Australian Capital Territory, Australia

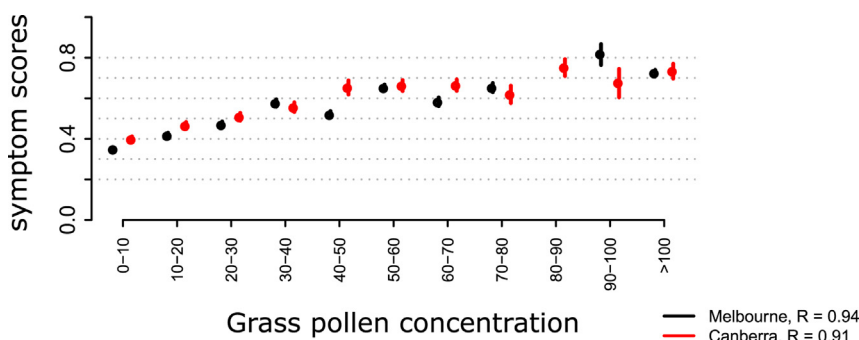
^f Western Sydney University and Campbelltown Hospital, New South Wales, Australia

^g School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

HIGHLIGHTS

- Citizen Science can help understand risk factors for allergic rhinitis (AR)
- Largest study of AR symptoms in an 'extra-clinical' population.
- Outdoor grass pollen levels are strongly correlated with increased AR symptom scores.
- User demographics also influence symptom scores.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 16 July 2019

Received in revised form 21 October 2019

Accepted 22 October 2019

Available online 22 November 2019

Keywords:

Hay Fever

Pollen

Allergic Rhinitis

Symptom Score

Smartphone

Citizen Science

ABSTRACT

Allergic Rhinitis (AR) affects over half a billion people worldwide with an estimated prevalence of 1 in 5 individuals in developed countries. Although ambient pollen exposure is a causal factor in AR, the symptom-exposure relationship is typically not studied in the broader community but in small, well-characterised cohorts drawn from clinical populations. To identify relationships between AR symptoms in the community and a range of environmental factors, we used a database containing over 96,000 symptom score reports collected over a 3-year period (2014–2016) through freely available smartphone apps released in two Australian cities, Melbourne and Canberra. Ambient pollen levels and symptom scores were strongly related, with grass pollen explaining most of the symptom variation. Other factors correlated with higher symptom scores included temperature ($R > 0.73$) and wind speed ($R > 0.75$). In general, worse symptom scores were reported by younger participants, women, and those who had taken medication for AR in the preceding 24 h. The strength of this relationship varied between the two cities. Smartphone-based symptom surveys offer a cost-effective means of studying real-world risk factors for AR in a broader 'extra-clinical' population.

© 2019 Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail address: Ed.L@unimelb.edu.au (E.R. Lampugnani).

1. Introduction

Allergic rhinitis (AR), or hay fever, has classic symptoms of rhinorrhoea, nasal congestion and sneezing, and commonly coexisting non-nasal symptoms such as conjunctival irritation. Globally, AR is estimated to affect one in five people (Kiotseridis et al., 2013) including up to 50% of adolescents (Bousquet et al., 2008). A major trigger of AR is the seasonal release of pollen by various plants, particularly grasses. AR can significantly impair an individuals' quality of life, interfere with daily activities and is associated with other complications including asthma exacerbations (Erbaş et al., 2012). Indeed, increases in the levels of airborne grass pollen have been positively correlated with increases in AR symptoms reported by patients with hay fever (Johnsen et al., 1992; Hanigan and Johnston, 2007; Erbaş et al., 2007; Johnston et al., 2009; Medek et al., 2012; Walter et al., 2019), as well as increases in hospital admissions for asthma (Silver et al., 2018). Across a population, reduced productivity and direct healthcare costs resulting from AR have a significant overall socio-economic impact (Meltzer and Bukstein, 2011).

As well as allergen exposure, known AR risk factors, as assessed in physician-managed cohorts, include environmental conditions and a range of demographic factors (Wang, 2005; Cazzoletti et al., 2015). The ever-increasing number of individuals with smartphones has led to a surge in demand for eHealth-related mobile applications (apps) that provide pollen information (Alquran et al., 2018; WHO, 2012). As eHealth apps also give researchers the ability to interact with these "extra-clinical" subjects - i.e. subjects not identified following a medical examination but self-identified based on symptoms - the use of app-based symptom surveys is becoming commonplace. Crowd-sourced symptom data from these large extra-clinical groups can contribute both to a better understanding of AR risk factors and have the potential benefit of improving the ability of participants to self-manage (Johnson et al., 2018).

AR is a major concern in Australia, as an estimated 20% of the population have AR and allergic and immune diseases are among the fastest growing chronic conditions in Australia (ASCI, 2013). Further, epidemics of thunderstorm asthma related to high grass pollen levels are well documented in South Eastern Australia (Marks et al., 2001; Silver et al., 2018). Notably, 20–30% of patients presenting with acute thunderstorm asthma attacks have histories of allergic rhinitis but lack a prior formal diagnosis of asthma (Thien et al., 2018).

The aim of this paper was to assess the utility of AR symptoms reported by the community for identifying factors contributing to these symptoms. We hypothesised that both demographic and environmental variables were associated with the severity of AR symptoms, given previous reports of their contribution in physician-managed cohorts. Data were gathered over three years (2014–2016) from Melbourne and Canberra, two major Australian population centres with significantly different geographies. During this period, Melbourne experienced a large-scale thunderstorm asthma event on November 21, 2016 (Thien et al., 2018) but survey results associated with this event were not specifically analysed as part of this paper.

2. Methods

2.1. Subject recruitment, demographic data and symptom score collection

Free apps for Android and iOS devices were developed and made available in Melbourne and Canberra in 2014–2016. The apps (called Melbourne Pollen and Canberra Pollen) were identical in design and active during the grass pollen seasons (October 1st to December 31st) when daily monitoring of airborne grass pollen

concentrations was taking place. The apps published daily pollen levels and forecasts and provided a quick and easy survey that asked users to rate their current hay fever symptoms on a severity scale with five categories, from no symptoms (1) to severe symptoms (5). Based on the Bousquet et al. (1994) "Quality of Life" scoring system, users were provided with the following guide on how to rate their symptoms:

- 1 = No Symptoms
- 2 = Mild symptoms (my symptoms were slight and a nuisance)
- 3 = Slight (my symptoms were obvious but tolerable)
- 4 = Moderate (my symptoms were obvious, inconvenient but still tolerable)
- 5 = Severe (my symptoms were obvious and intolerable)

Individuals self-selected themselves for participation by interacting with the survey widget and only the first survey score recorded by a device in a 24 hr period was used. Data collected was approved for analysis by the Human Research Ethics Committees of the University of Melbourne and Melbourne Health (1647764.1 and QA2015148). Although subjects were not identifiable, devices were assigned a universally unique identifier (UDID) so that submissions from the same device could be linked. Users could provide a survey response at any time of day, at which point the time and device UDID were also recorded and users could elect to provide location (GPS) data as well. Further, after submitting their symptom scores, users were prompted with the yes/no question "Have you taken any hay fever medication today?"

In 2015 and 2016 the apps provided an additional survey that asked users for the following demographic information:

- What is your age group?
- What is your biological sex?
- What is your occupation?
- In an average week how many hours do you spend outside?
- Have you been diagnosed with asthma by a doctor?
- Have you been diagnosed with allergic rhinitis (hay fever) by a doctor?
- Are you taking any regular preventative treatments for your hay fever?

Outcome variables are described in Supplemental Table 1.

As described in the results section, the dataset was filtered to include only those survey results recorded between October 1 and December 31 with GPS coordinates that were within 50 km of the Melbourne trap site or 20 km of the Canberra trap site (the rationale for these thresholds is outlined below). A summary of the number of response examined in this study after filtering is provided in Table 2.

2.2. Study sites and airborne pollen collection

Melbourne Australia has a population of approximately 4.9 million and is located on the south east coast of mainland Australia (Fig. 1). The city has a largely rural landscape to the north and west, ocean to the south and forested areas to the east. Canberra, Australia's capital city, has around 0.5 million inhabitants and is landlocked. The surrounding countryside is a matrix of farmland and natural forest.

Daily grass and total pollen concentrations were measured using a Burkard volumetric trap (Burkard Manufacturing Co. Ltd., Rickmansworth, Hertfordshire, UK) as previously described (Ong et al., 1995). The Melbourne sampling site was located at the University of Melbourne 1.7 km north of central Melbourne at coordinates -37.7971 and 144.9648 (Fig. 1). The Canberra sampling site was located at the Australian National University at coor-

Table 1

Summary of pollen recorded for Melbourne and Canberra (2014–2016) grass pollen seasons (Oct 1 – Dec 31).

	Melbourne						Canberra					
	2014		2015		2016		2014		2015		2016	
	Grass	Other	Grass	Other	Grass	Other	Grass	Other	Grass	Other	Grass	Other
Mean count	13.02	79.03	11.1	57.1	28.9	101.36	26.66	36.02	32.19	50.41	15.08	24.04
Median count	9	52.5	5	32.5	17	58.5	14	25	26.5	40	9	9
SD count	13.4	66.75	13.69	73.01	33.78	115.39	29.51	31.96	25.81	38.01	15.68	44.89
Number of days reporting	92	92	92	92	92	92	91	91	90	90	92	92
Max count	54	338	58	418	154	714	130	126	115	234	69	313
Total	1,198	7,271	1,021	5,253	2,659	9,325	2,426	3,278	2,897	4,537	1,387	2,212
Number of lows	68	–	72	–	51	–	50	–	32	–	64	–
Number of mediums	22	–	17	–	21	–	24	–	37	–	21	–
Number of highs	2	–	3	–	20	–	14	–	19	–	7	–
Number of extremes	0	–	0	–	6	–	3	–	2	–	0	–

Table 2

Numbers of submissions and unique devices for Melbourne and Canberra for 2014–2016 after filtering.

Filtered Data	Melbourne				Canberra			
	2014	2015	2016	Total	2014	2015	2016	Total
Number of responses from Android devices	3525	729	7926	12180	2504	790	2422	5716
Number of responses from iOS devices	5051	2980	21876	29907	3314	2894	6298	12506
Total number of responses	8576	3709	29802	42087	5818	3684	8720	18222
Number of unique Devices	2420	1379	6226	9475	1213	921	1370	3105
Mean number of submission per device	3.5	2.6	4.7	4.4	4.7	4.0	6.3	5.8
Median number of submission per device	2	1	2	2	2	1	3	2
Maximum number of submission per device	59	53	74	122	67	67	69	110
Percentage of devices with a single submission	46.3	61.4	39.0	42.7	39.0	51.4	30.2	37.3
Mean number of submissions per device among multiple submitters.	5.7	5.3	7.2	7.0	7.2	7.1	8.6	8.7

The “full year” spans Jan 1 to Dec 31, the “grass season” is defined as Oct 1 to Dec 31, and the “filtered data” are restricted to the grass season only and to within 50 km of the trap location for Melbourne and within 20 km for Canberra.

dinates –35.1632 and 149.7100. Samples were collected at 16:00 h (local time) with the measured concentration representing the 24-hour average concentration from 16:00 h the previous day in pollen grains/m³ of air. The measured concentration is associated with the day the measurement was made, as this day covers 67% of the measurement period. Daily grass pollen concentrations are described within this manuscript in terms of categorical levels (following Ong et al., 1995): ‘low’ (an average concentration of 0–19 grains/m³), ‘moderate’ (20–49 grains/m³), ‘high’ (50–99 grains/m³) and ‘extreme’ (≥100 grains/m³). Additional attributes of the study sites, namely population density, mean annual rainfall and topography are presented in Supplemental Figs. 1 and 2.

2.3. Environmental and weather variables

Daily temperature (Celsius), relative humidity, rainfall and wind speed records for Melbourne (the Olympic Park automatic weather station) and Canberra (Canberra airport) were provided by the Australian Bureau of Meteorology. Averaged or integrated versions of these weather variables were used, rather than the instantaneous values. For temperature and relative humidity, the average daily value was used, and for rainfall, the sum from 00:00 h to 23:59 h (local time) was used. For wind speed, the average wind speed and direction was calculated based on averages of the north–south and east–west components of the hourly wind vectors.

Air quality data for Canberra was provided by ACT Health through the Open Data Portal (www.data.act.gov.au) and for Melbourne by the Victorian Environmental Protection Agency under a data request and is now also available through the Open Data Portal. To estimate the daily particulate matter (PM) concentration in each city, concentrations were averaged across three monitoring sites in Canberra (PM_{2.5}) and 11 sites in Melbourne (PM₁₀). While it would have been preferable to use PM_{2.5} data for both locations,

there were more PM₁₀ sites operational across Melbourne during the study period and the temporal coverage was much better.

2.4. Statistical analysis

The R statistical computing environment was used for all computations (R Core Team, 2015).

3. Results

3.1. Melbourne and Canberra grass pollen seasons 2014–2016

Table 1 shows summary measures of the 2014–2016 Melbourne and Canberra grass pollen seasons (October 1 to December 31 each year). Average daily grass pollen concentrations varied at the two locations and over the three seasons, being on average 1.4 times higher in Canberra, and for a given site the seasonal averages ranged by a factor of 2.1–2.6 (Fig. 2 and Supplemental Fig. 3). There were five ‘high’ (50–99 pollen grains/m³) and no ‘extreme’ (≥100 pollen grains/m³) grass pollen days in Melbourne in the 2014 and 2015 seasons, and 20 ‘high’ grass and 6 ‘extreme’ grass pollen days in 2016. The 2016 season was also notable for the November 21 thunderstorm asthma event in which grass pollen levels likely played a critical role (Thien et al., 2018). The 2014 and 2015 Canberra seasons were heavier in terms of the total amount of grass pollen recorded (2–2.9 fold more than for the corresponding Melbourne season) and number of ‘high’ or ‘extreme’ days (17 and 21, respectively). By contrast, 2016 was proportionally a lighter year with only 7 high grass pollen count days.

Daily concentrations of non-grass pollen (all pollen other than grass pollen) were in general much higher than grass pollen concentrations, being on average 4.5 times higher in Melbourne and 1.5 times higher in Canberra (Table 1). Consistent with previous

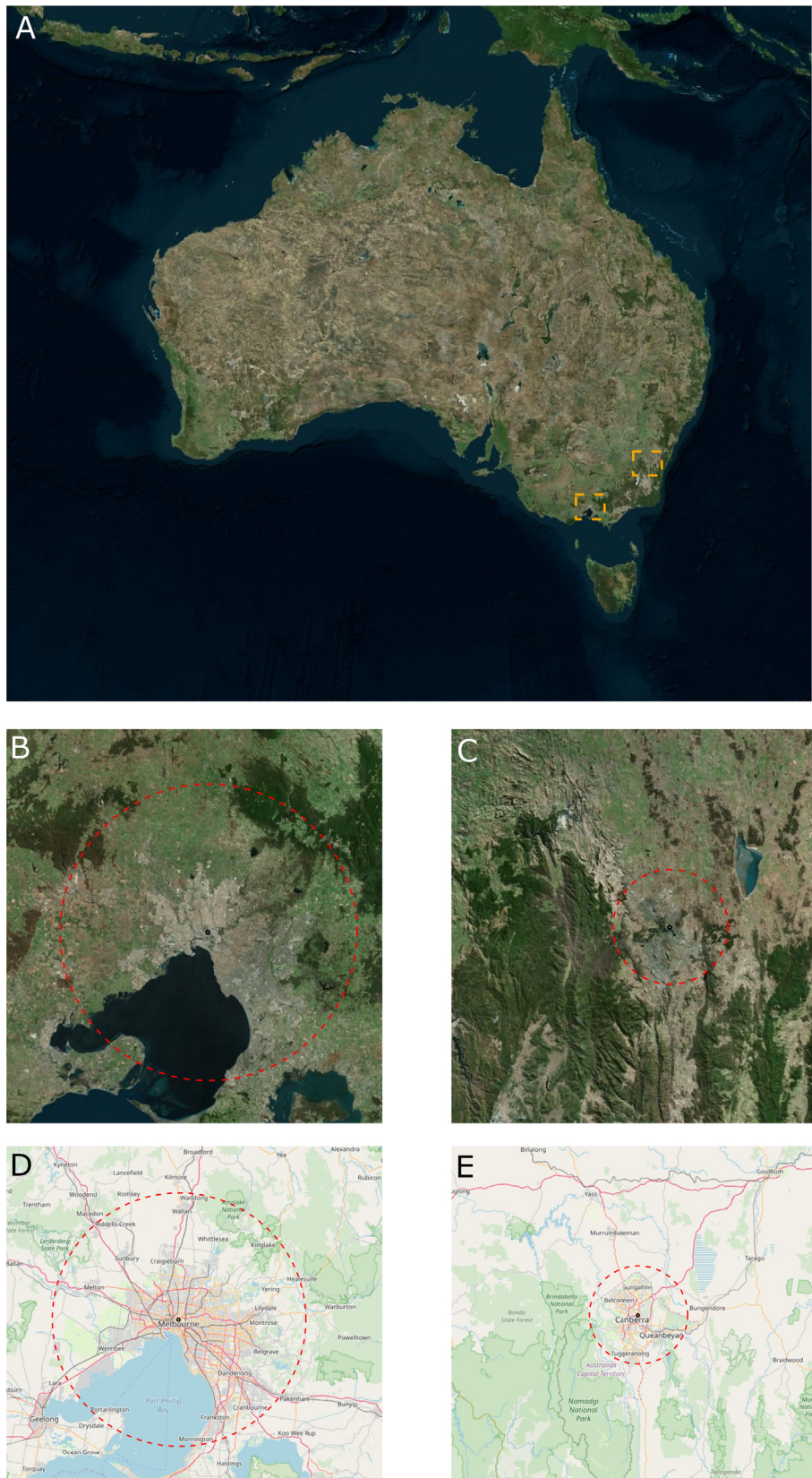


Fig. 1. Locations of the study regions within Australia. (A) Map of Australia showing the locations of the study regions within Australia. (B-C): Detailed satellite imagery and (C-E) localities of (B,D) Melbourne and (C,E) Canberra. Circles are shown around the trap site depicting 50 km for Melbourne and 20 km for Canberra. Imagery provided by Bing or OpenStreetMap.

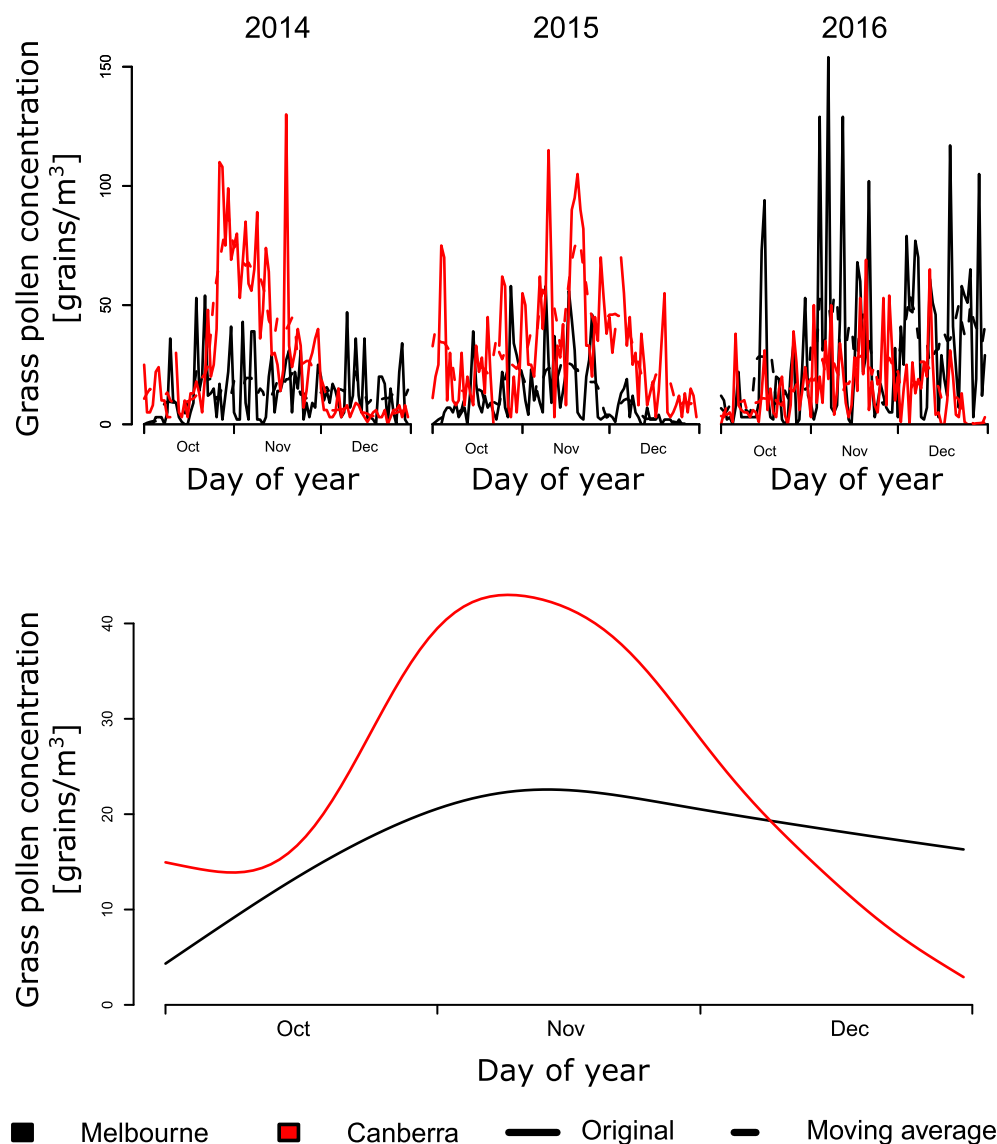


Fig. 2. Time-series of the grass pollen concentrations across seasons. Grass pollen concentrations in Melbourne (black) and Canberra (red). The upper two rows show the raw counts (solid line) and seven-day moving averages (dashed line). The bottom panel shows a smoothed version, based on all four years of data. Smoothing was performed using a generalised additive model with a spline fit for the one predictor variable (the day of the year). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

studies (Haberle et al., 2014), non-grass pollen levels declined over the October–December period at both sites (Supplemental Fig. 3). The ‘non-grass’ pollen fraction is taxonomically very broad and its composition differs between Canberra and Melbourne. The Melbourne site is characterised by large amounts of Cupressaceae (cypress) pollen (around 60% of the total) and smaller amounts of native Myrtaceae (eucalypt) pollen, as well as pollen from introduced northern hemisphere trees like Oleaceae (the ash, *Fraxinus*) and *Betula* (birch) (Haberle et al., 2014). The Canberra non-grass pollen fraction contains much less Cupressaceae pollen than Melbourne but has proportionally more *Pinus* (pine) and *Rumex* (sorrel) pollen (Haberle et al., 2014).

3.2. App usage and filtering of survey responses

Free Melbourne and Canberra pollen smartphone apps were released in time for the 2014 season and were updated annually. By the end of 2016, the Melbourne Pollen app had been installed on 49,957 devices and utilised 547,137 times and the Canberra Pol-

len app installed on 15,146 devices and used 134,200 times (Supplemental Table 2). A simple survey on the apps allowed users to rate their daily AR symptoms on a 5-point scale. Supplemental Table 2 shows that 15,132 devices reported 67,260 survey responses using the Melbourne Pollen app (75% with a GPS location) and 4,869 devices reported 28,917 responses using the Canberra Pollen app (74% with a GPS location). The number of submissions per device exhibited a power-law decay (Fig. 3A), with the highest number of reports per device over the three-year period being 144 for Melbourne and 139 for Canberra (Supplemental Table 2). Approximately 44% of devices across the two cities recorded only a single survey response (Supplemental Table 2).

Three filters were applied to the dataset of survey responses. First, only those survey responses with associated GPS coordinates were used. Second, as the frequency of daily symptom reporting was highest when the apps were providing information about grass pollen (Fig. 3B), the data was filtered to include only those survey responses received in the busy October 1–December 31 period. During this time there was on average 315 daily submissions

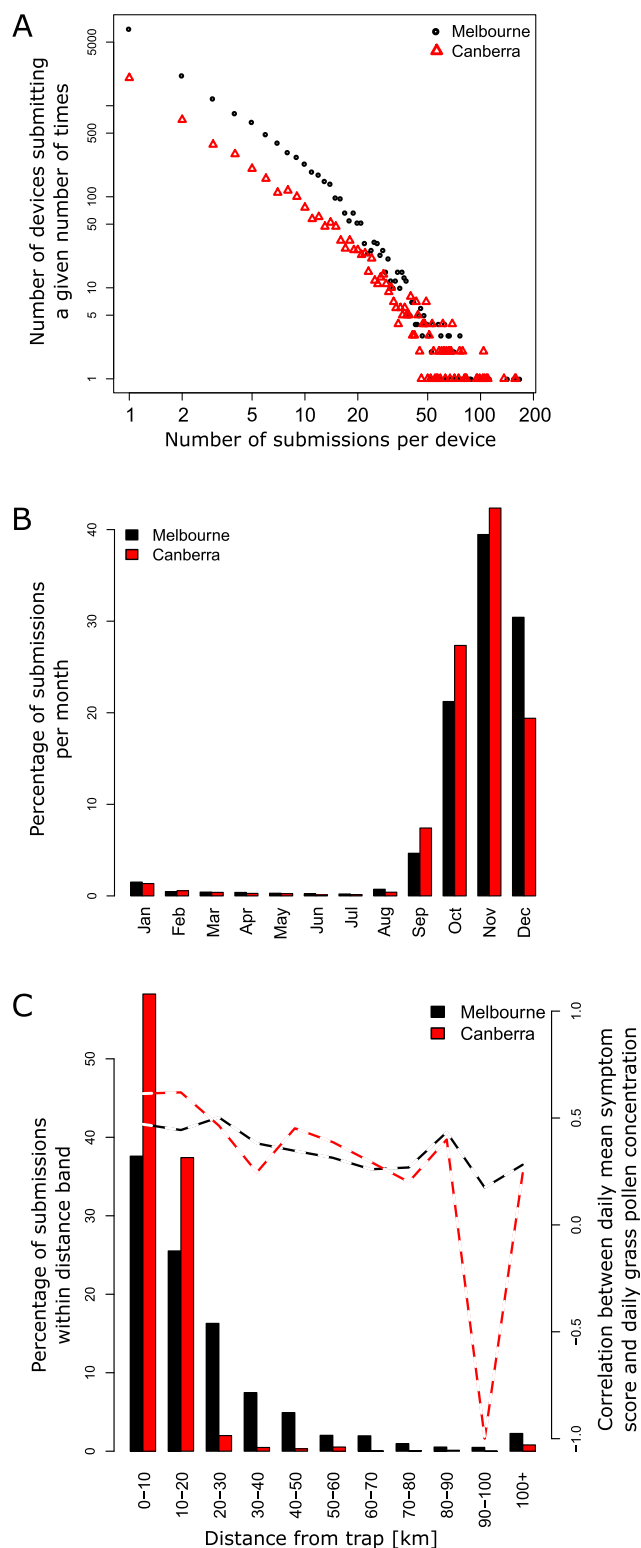


Fig. 3. Number of survey submissions. (A) Number of submissions per device. (B) Percentage of responses per month of year. (C) Number of responses as a function of distance from trap location (bars), and the correlation between daily symptom scores and daily grass pollen at different separation distances (dashed lines).

over the two sites, with the daily maximum being 1,622 submissions. The number of reports fell to < 10 per day for the rest of the year. Finally, a small number of submissions were received from hundreds of kilometres from the pollen monitoring sites, including some from other countries. To detect meaningful rela-

tionships between local environmental conditions and conditions to which users were exposed, the analysis was further restricted to include only those surveys received from a device within 50 km of the Melbourne trap and 20 km of the Canberra trap (Fig. 3C). This criterion is consistent with Canberra being a smaller city than Melbourne (Fig. 1A-B) and with the majority of the population being < 100 km from the city centre (see Supplemental Fig. 1A and B).

3.3. Environmental factors and their influence on AR symptom scores

Much of the statistical analysis was performed using a dichotomised variable (i.e. whether the symptom score was ≥ 3), rather than the original five-point scale; this was done to simplify the analysis and allowing readily-interpretable confidence intervals for this binomial proportion. An initial question was how much data-loss occurred because of this dichotomisation. Clearly for individual data points there is considerable data loss (e.g., once dichotomised, 3 is indistinguishable from 5). However, across a group of scores much information was retained: examining all data for a given day, there was a very clear relationship ($R \geq 0.95$ across both cities) between the mean of the symptom scores and the fraction of symptom scores ≥ 3 (called F3; Supplemental Fig. 4B). Symptom scores of 1 and 2 comprised 51% of all submissions, thus this threshold was the closest to the median, making F3 the most informative dichotomisation. Other versions of the fractional variable (e.g. F2, whether symptom scores were ≥ 2 , or F5, whether symptom scores were ≥ 5) returned lower R values than F3 (Supplemental Fig. 4A, C and D). For clarity, F3 is defined as:

$$F3 = (\text{number of symptom survey submissions with values } \geq 3) / (\text{total number of submissions})$$

Fig. 4A-B summarises relationships between user symptom scores and various temporal factors. AR symptoms were worse on Sundays and Mondays, an observation that is consistent with Melbourne's weekly cycle of asthma-related hospital admissions (Silver et al., 2018). For example, F3 was 0.123 (0.068) higher on Sundays than Wednesdays in Melbourne (Canberra) with the null hypothesis for the χ^2 -test of equality of proportions rejected at $p < 10^{-10}$ ($p < 10^{-3}$). Users in Melbourne also reported symptoms were worse in the early morning (e.g., 3–4 am) and were better in the evening (around 7 pm). The F3 at 4am was 0.25 (0.11) higher in Melbourne (Canberra) with χ^2 -test result $p < 10^{-10}$ ($p < 10^{-10}$). A range of hypotheses for this 3–4 am increase in symptoms can be considered, such as house dust mite allergies in co-sensitised individuals or non-allergic hyperactivity associated with the nighttime temperature drop, but the available data does not allow these and other plausible hypotheses to be tested. This diurnal pattern was less obvious in Canberra, presumably because the smaller number of symptom scores submitted between 12am–6am (Supplemental Fig. 5) made the pattern harder to infer.

The apps utilised in this study provided users with a forecast of the expected grass pollen levels in the coming 24 h. These forecasts were issued at 4 pm each day during the study period and could potentially influence user survey scores. Considering all of the data, there was a rise in the number of responses submitted at around 4 pm, but notably there was no change in reported AR symptoms at either site (Fig. 4B, Supplemental Fig. 5B). While this feature of the survey could potentially influence the relationship between AR symptom scores and grass pollen levels, it would not affect the relationship with other environmental and demographic variables and medication usage.

Fig. 3C shows that the correlation between average daily symptom scores and grass pollen counts decayed with distance from trap. A Williams' test (Williams, 1959; Steiger, 1980), which tests for differences between a pair of correlations where the correlations for both relate to a common variable, was used to analyse

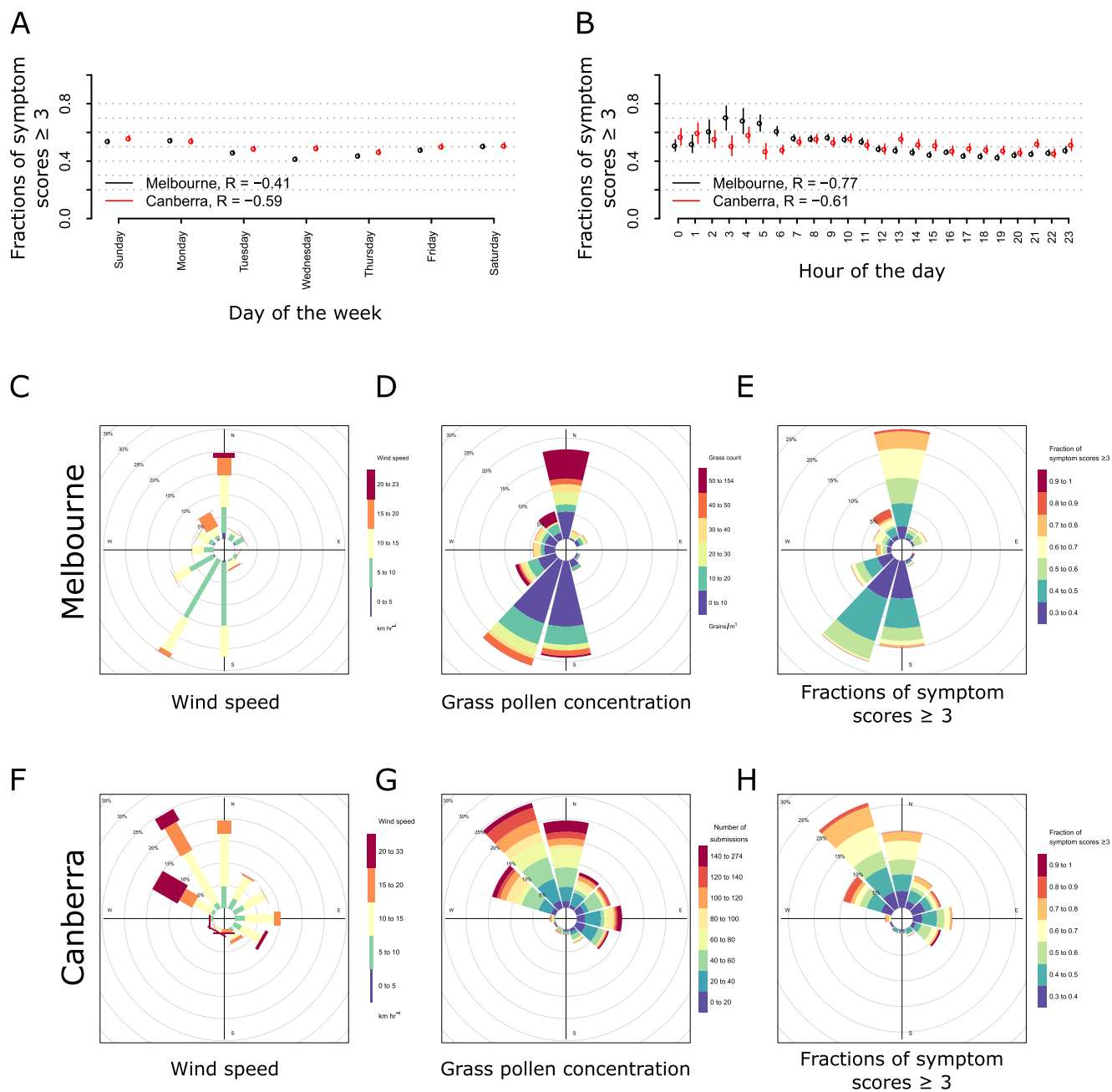


Fig. 4. User symptom scores and various spatial and temporal factors. (A–B) Fraction of symptom scores ≥ 3 (F3) as a function of day-of-week and time-of-day. Vertical bars show a 95% confidence interval for the corresponding fractions. (C,F) Distribution of wind speeds per wind sector. (D,G) Distribution of grass counts per wind sector. (E, H) Distribution of the fraction of survey scores ≥ 3 per wind sector. The wind-sector plots for Melbourne appear in C to E, and for Canberra F to H.

correlations between grass pollen levels and survey scores recorded at different distance bands from the pollen trap. The first correlation compared grass pollen levels with symptom scores within the 0 to 10 km band, and the second correlation compared the grass pollen levels with symptom scores in the x to $(x + 10)$ km band (for x in 10, 20, ..., 90); the common variable here was the grass pollen level recorded at the trap site. The test accounts for the sample sizes underlying the correlations and the correlation between symptom scores in the two bands. Differences between the 0–10 km correlation and correlations at greater distances became statistically significant (using Williams' test and $p = 0.05$ significance threshold) beyond 50 km (20 km) away from the Melbourne (Canberra) spore trap. This analysis also informed the distance thresholds used in the third data filtering step (see Methods).

The relationship between grass pollen and wind direction and strength was examined to determine if the user's position relative to the spore trap influenced the score (Fig. 4C–H). In Melbourne, winds during the October–December period were predominantly from the North, South and West; the Canberra site experienced winds mostly from the North-West, North, East and South-East. In Melbourne, the highest grass pollen levels and highest average symptom survey scores coincided with winds from the North (northerly winds were experienced on 52% of days when the F3 exceeded 0.5); conversely, South and South-West winds were associated with lower grass pollen levels and lower symptom scores (winds from the southerly quadrant were experienced on 62% of days when the F3 was below 0.5). In Canberra, high symptom scores and grass pollen levels were less concentrated within speci-

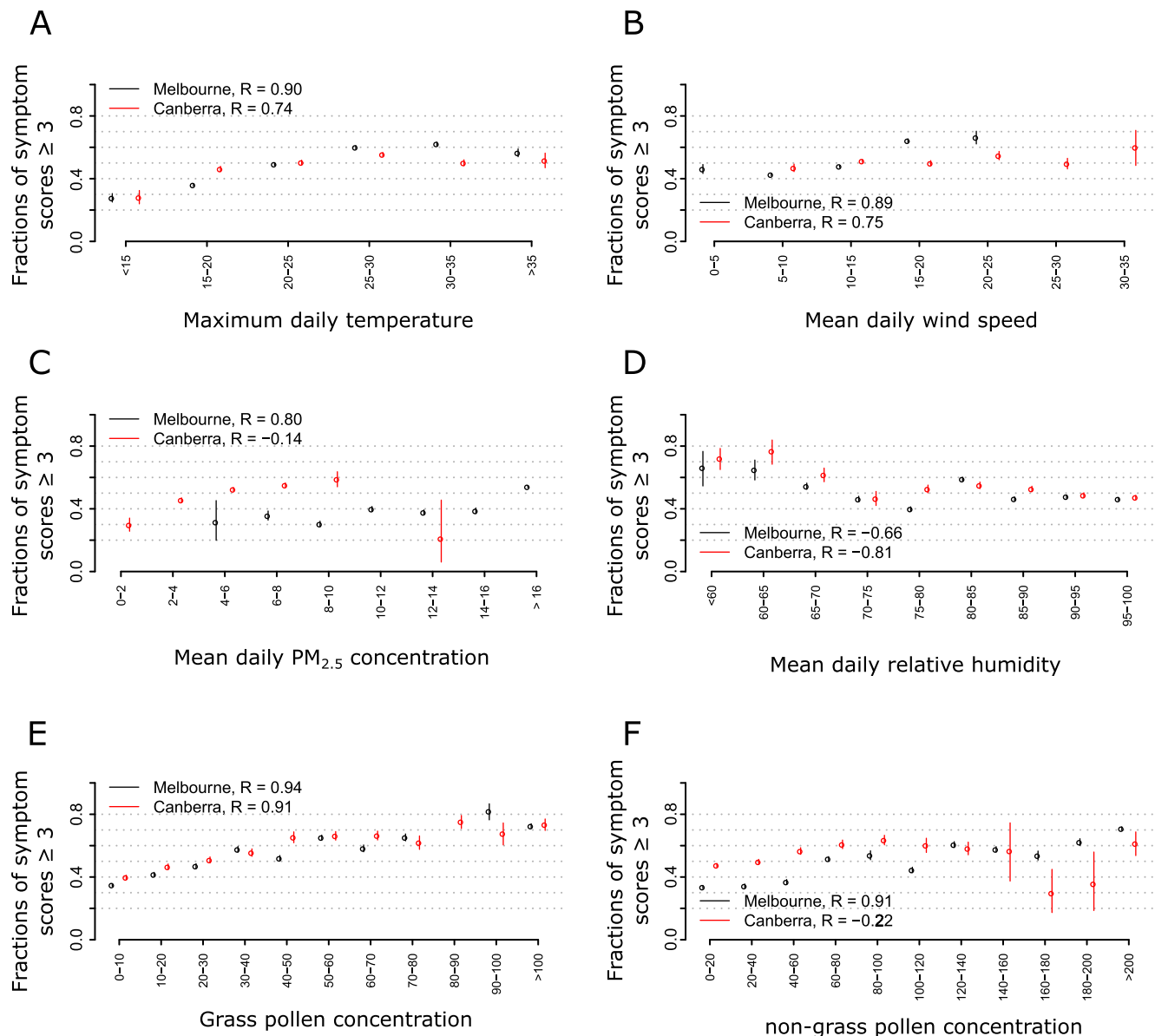


Fig. 5. Marginal relationships between user symptom scores and various environmental factors. Vertical axes show the fraction of symptom scores ≥ 3 (F3) plotted against the indicated environmental factor, with Canberra data shown in red and Melbourne data shown in black. R values for the Pearson correlation coefficients are shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fic wind sectors; this likely reflects the fact that agricultural land can be found within 30 km of the city in most directions. The number of responses obtained was also influenced by the bearing relative to the spore trap (Supplemental Fig. 5C-D). At both sites, there was a weak relationship between the number of responses per day and mean symptom score ($R = 0.25$ for Melbourne, $R = 0.37$ for Canberra; Supplemental Fig. 5E).

Fig. 5 shows the relationships between several variables (wind, temperature, particulate matter, pollen, humidity) and symptom scores in Canberra and Melbourne. A strong correlation exists between the user symptom score and the level of grass pollen in the air (Pearson correlation coefficient of $R = 0.94$ and 0.91 for Melbourne and Canberra, respectively). This relationship appears close to linear over the range $0-100$ grains/ m^3 , and given that the F3 cannot exceed 1.0, this relationship must be non-linear over a wider range; we have not explored such non-linearity, but within this range the slope and intercept of the linear regression are not

significantly different for the two sites. Although in Melbourne, F3 and levels of non-grass pollen were also highly positively correlated ($R = 0.91$), the two variables were negatively correlated in Canberra (-0.22). Higher symptom scores showed an increasing relationship with temperature ($R > 0.73$) and wind speed ($R > 0.75$) and were negatively correlated with relative humidity ($R < -0.66$). The relationship between the number of responses and associated environmental conditions was mainly explained by the frequency of the environmental conditions; for example, the percentage of responses on days above 35 degrees Celsius was very small, largely because there were very few such days (Supplemental Fig. 6). However, given the association between number of responses and the magnitude of the symptoms reported, fewer submissions were made on days with low grass pollen than would be expected based on the frequency of these days alone, and conversely for high grass pollen days (Supplemental Fig. 6).

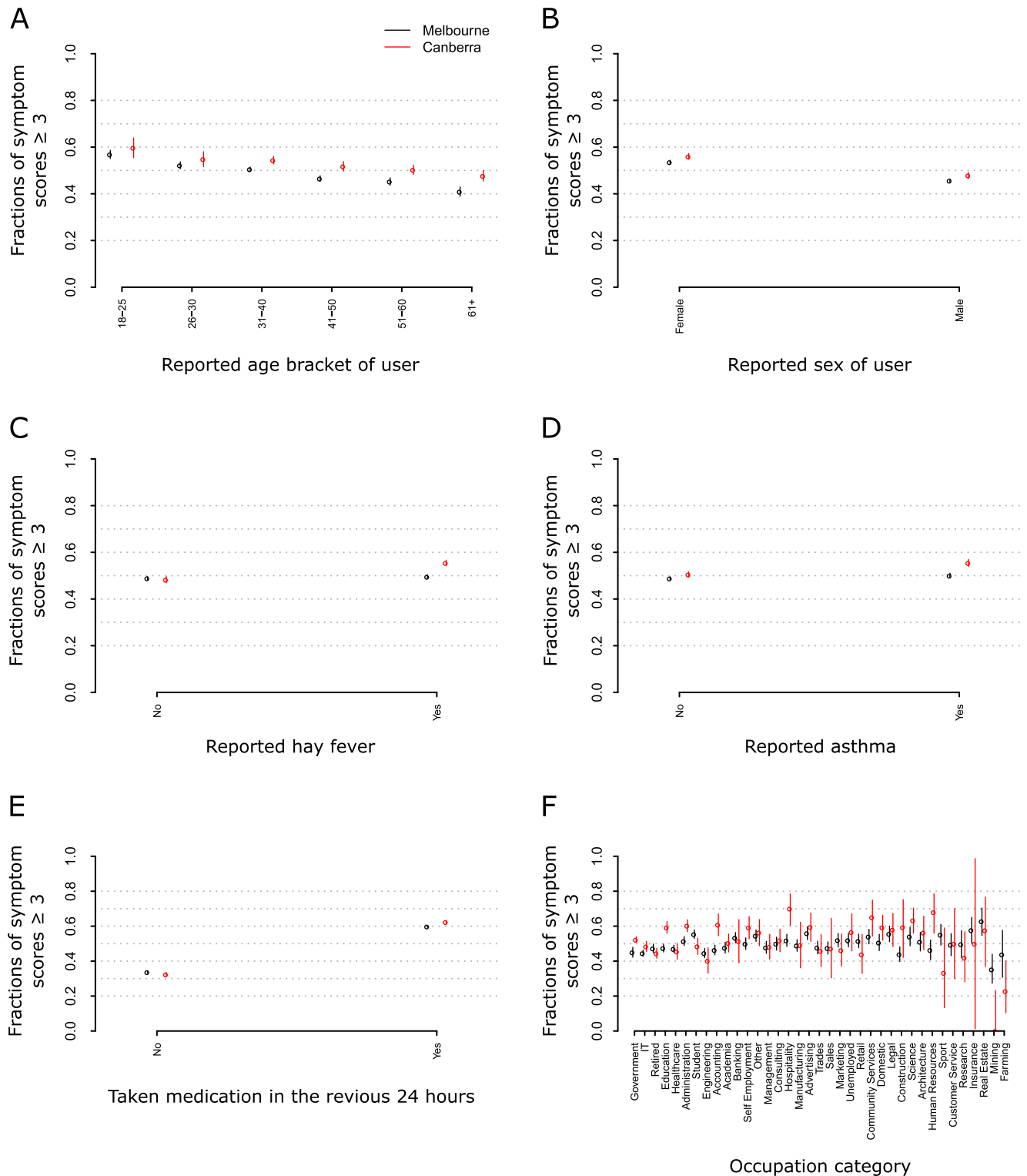


Fig. 6. Relationships between user symptom scores and various demographic factors. Vertical axes show the fraction of symptom scores ≥ 3 (F3) plotted against the indicated demographic factor, with Canberra data shown in red and Melbourne data shown in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. User demographic factors and their influence on AR symptom scores

In 2015 and 2016 app users could elect to provide additional information about themselves and how they managed their AR through an extended in-app survey. Demographic data was pro-

vided by 52% of unique devices (56% of submissions) in Melbourne, and 55% of unique devices (59% of submissions) in Canberra. There was an approximately equal proportion of men and women, with women accounting for 47% of users in Melbourne and 52% of users in Canberra. People 31–40 years of age provided the largest proportion of submissions and users (30–35% for both sites), with peo-

ple aged 41–50 being the next largest (about 20% of users and submissions at both sites). This is consistent with smartphone ownership statistics for Australia, which indicate that younger people are generally more likely to own a device than those in older age brackets (Deloitte 2017). Considering the relationships between user symptom scores and these variables, worse symptom scores were reported by younger participants, women, and those who had taken medication for AR in the preceding 24 h (Fig. 6). The self-reported occupation data segregated users into many categories, most with very few submissions (Supplemental Figure 7). Given the high degree of stratification of the sample there were few obvious relationships between professions and higher or lower survey symptom scores.

4. Discussion

This paper analyses a large number of app-based AR symptom scores collected from 2014 to 2016 around Canberra and Melbourne, two major cities in south-eastern Australia. Key to the high level of participation in this study was the rapid public uptake and sustained engagement with the two apps. This uptake occurred organically and without the need to advertise for participants. Although not the first app-based survey of AR, with nearly 50,000 participants this was by far the largest study of AR symptoms in Australia. Johnston et al. (2018) reported results of a similar survey in Tasmania but with fewer downloads (<3,500) and a smaller number of symptom reports. The number of participants in this study is also far larger than those reported in similar studies in Europe. For instance, Karatzas et al. (2014) and de Weger et al. (2014) utilised an online patient hay-fever diary which users could access through a web interface and had 1,600 and approximately 500 participants, respectively. While the publicity associated with the 2016 thunderstorm asthma event played some role in recruitment, Table 2 shows that this event was not the only driver of engagement, as large numbers of symptom scores were reported in the preceding years and in Canberra as well as Melbourne.

Rather than a consequence of specific events, a key factor contributing to the high level of participation in the survey was its simplicity. The symptom surveys described in Johnston et al. (2018) and other similar studies (e.g., Bastl et al., 2014; Karatzas et al., 2014; de Weger et al., 2014; Werchan et al., 2018) are lengthier and more detailed than the one presented here. Because the frequency of user interactions with mobile health apps is directly affected by the ease of interaction, the Melbourne and Canberra pollen apps used a 'single breath' survey design able to be completed in a couple of screen touches. Lower levels of app engagement are seen when users are required to provide detailed information or where data entry requires more effort (Choi and Tulu, 2017). Despite its simplicity, however, repeat engagement with the apps was low and the number of submissions per device exhibited a power-law decay (Fig. 3A). This low level of repeated engagement with the app and low number of submissions per device is consistent with observations reported in other similar studies (Karatzas et al., 2014; de Weger et al., 2014; Werchan et al., 2018).

A second factor contributing to the success of this study is the well-defined and relatively short AR seasons of Melbourne and Canberra. These seasons have previously been attributed to airborne grass pollen, which is generally considered Australia's worst outdoor allergen source (Schäppi et al., 1998). While earlier studies have shown correlations between levels of ambient grass pollen and measures of AR symptoms in Canberra and Melbourne (Schäppi et al., 1998; Medek et al., 2012), these studies were small in scale, involving dozens of participants, and restricted to participants with a confirmed grass pollen allergy. Although users of the

Melbourne and Canberra pollen apps were not known to have a diagnosed pollen allergy, the present study was still able to confirm the overwhelming importance of grass pollen as the likely cause of allergies during the late austral spring and early summer. A similar strong association between grass pollen levels and asthma admissions to hospitals has previously been found in Melbourne and has helped to define a high risk spring asthma season (Erbaş et al., 2012; Silver et al., 2018). The strong linear relationship between ambient grass pollen levels and levels of AR up to an average daily concentration of 100 grass pollen grains m^{-3} (Fig. 5) was similar across the two study sites. Overall, this relationship was stronger and less complex than that between symptom scores and birch, grass and ragweed pollen levels in Austria and Germany (Bastl et al., 2014). The correlation with symptom scores decreased with distance from the monitoring site, suggesting either a radius around which the central site is representative or spatial clustering of the population around this site, with the smaller number of reports at greater distances being less representative.

While grass pollen was the major factor of seasonal AR in Melbourne and Canberra, it is likely that other factors also contribute. For instance, symptom scores were also associated with the non-grass pollen fraction, although the shape of this response varied between the two cities in keeping with differences in the composition of the 'non-grass' pollen fractions of Canberra and Melbourne (Haberle et al., 2014). An extended pollen monitoring season and a finer resolution taxonomic breakdown of the non-grass pollen fraction can potentially highlight further associations with AR symptoms. However, a limitation of this approach is that, unlike grasses, which are grown over large agricultural areas, most introduced species in Australia are less widespread in their distribution and their pollen more likely to show spatial variation in concentration. Additionally, concentrations of some common and typically indoor allergen sources, such as fungal spores and dust mites, are hard to determine on a spatial scale representative of a city and other AR triggers are not environmental in origin (Bousquet et al., 2001). Thus, although the contribution of grass pollen was evident in the survey data, other factors, such as strong spatial heterogeneity in the distribution of allergenic species, may limit the utility of the methods described here to other situations.

Smartphone-based symptom surveys offer a cost-effective means of obtaining large sample sizes and thus improved statistical power. Furthermore, by allowing access to a wider range of participants and potentially a broader range of disease severities, such datasets may better reflect disease burden in the community (Tan et al., 2018). Two limitations of this study, however, were the self-selecting nature of the population sample and the potential for bidirectional causality, in that access to grass pollen forecasts may have influenced reported AR symptoms; our findings should thus be cross-checked through further research using prospective cohorts or sampling studies. However, many of the features of the study group are consistent with previous findings, such as men and older adults reporting less severe AR symptoms (Olsson et al., 2003; Nathan et al., 1997). Further, the worst average hay-fever symptoms across the week were reported on Sundays and Mondays, echoing results from a recent study of asthma-related hospital admissions in Melbourne (Silver et al., 2018). The increase in hay-fever symptoms on Sundays and Mondays may simply be because of behavioural changes on weekends affecting time spent outside and thus exposure to outdoor aeroallergens. The strong association between daily grass pollen concentrations and hay fever symptoms apparent in both cities has been reported elsewhere, as have associations between grass pollen and asthma (e.g. Erbaş et al., 2012; Huynh, 2010; Annesi-Maesano et al., 2012).

Future studies using this dataset will examine the utility of the relationships described here to improve the accuracy of daily grass

pollen forecasts in Canberra and Melbourne and their ability to provide personalised predictions of next-day AR symptoms.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the contribution to this paper made by users of the Melbourne and Canberra Pollen Count apps. Air quality data was provided by the Victorian Environment Protection Authority and we thank Michael Stringer and Jason Choi for their help with obtaining it. Weather data was provided by the Bureau of Meteorology and we thank Tony Bannister for his help with obtaining it. Technical support for app development and data storage was provided by the University of Melbourne, especially Usha Nattala, Uli Felzmann, Joshua Lilly, Chris Stroud, Robert Sturrock and Baybars Tetik. Funding for the Melbourne Pollen Monitoring Program and for analysis was provided by the University of Melbourne. Funding for the Canberra Pollen Monitoring Program was provided by the Australian National University and ACT Health. We thank Pamela Burton (Campbelltown Hospital and Western Sydney University) for her assistance with obtaining ethics approval and colleagues in the AusPollen network (NHMRC grant GNT1116107) for ongoing discussions.

Author contributions

ERL, JDS and EJN conceived and designed the experiments and ERL built the mobile apps that provided the user symptom data. SH and EJN provided pollen data. JDS, ERL and KS analysed the data with input from EJN. All co-authors contributed to writing the manuscript and approved the final draft.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2019.135147>.

References

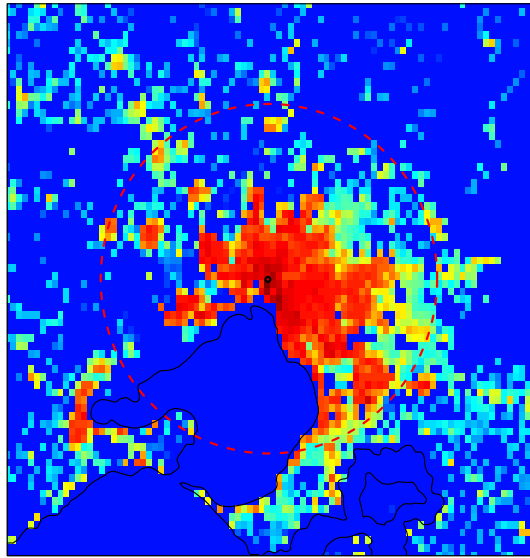
- Annesi-Maesano, I., Rouve, S., Desqueyroux, H., Jankovski, R., Klossek, J.M., Thibaudon, M., Demoly, P., Didier, A., 2012. Grass pollen counts, air pollution levels and allergic rhinitis severity. *Int. Arch. Allergy Immunol.* 158 (4), 397–404.
- ASCIA (2013). Allergy and Immune Diseases in Australia (AIDA) Report 2013. https://allergy.org.au/images/stories/reports/ASCIA_AIDA_Report_2013.pdf
- Alquran, A., Lambert, A.K., Farouque, A., Holland, A., Davies, J., Lampugnani, E.R., Erbas, B., 2018. Smartphone Applications for Encouraging Asthma Self-Management in Adolescents: A Systematic Review. *Int J Environ Res Public Health* 15 (11). <https://doi.org/10.3390/ijerph15112403>.
- Bastl, K., Kment, M., Jäger, S., Bergmann, K.C., Berger, U., 2014. Development of a symptom load index: enabling temporal and regional pollen season comparisons and pointing out the need for personalized pollen information. *Aerobiologia* 30, 269–280.
- Bousquet, J., Knani, J., Dhivert, H., Richard, A.L.A.I.N., Chicoye, A.N.N.I.E., Ware, J.E., Michel, F.B., 1994. Quality of life in asthma. I. Internal consistency and validity of the SF-36 questionnaire. *Am. J. Respir. Crit. Care Med.* 149 (2), 371–375.
- Bousquet, J., van Cauwenberge, P., Khaltaev, N., Aria Workshop Group, World Health Organization, 2001. Allergic rhinitis and its impact on asthma. *J. Allergy Clinical Immunol.* 108 (Supplement), S147–S334.
- Bousquet, J., Khaltaev, N., Cruz, A.A., Denburg, J., Fokkens, W.J., Togias, A., Zuberbier, T., Baena-Cagnani, C.E., Canonica, G.W., van Weel, C., Agache, I., Ait-Khaled, N., Bachert, C., Blais, M.S., Bonini, S., Boulet, L.P., Bousquet, P.J., Camargos, P., Carlsen, K.H., Chen, Y., Custovic, A., Dahl, R., Demoly, P., Douagui, H., Durham, S. R., van Wijk, R.G., Kalayci, O., Kaliner, M.A., Kim, Y.Y., Kowalski, M.L., Kuna, P., Le, L.T., Lemiere, C., Li, J., Lockey, R.F., Mavale-Manuel, S., Meltzer, E.O., Mohammad, Y., Mullol, J., Naclerio, R., O'Hehir, R.E., Ohta, K., Ouedraogo, S., Palkonen, S., Papadopoulos, N., Passalacqua, G., Pawankar, R., Popov, T.A., Rabe, K.F., Rosado-
- Pinto, J., Scadding, G.K., Simons, F.E., Toskala, E., Valovirta, E., van Cauwenberge, P., Wang, D.Y., Wickman, M., Yawn, B.P., Yorgancioglu, A., Yusuf, O.M., Zar, H., Annesi-Maesano, I., Bateman, E.D., Ben Kheder, A., Boake, D.A., Bouchard, J., Burney, P., Busse, W.W., Chan-Yeung, M., Chavannes, N.H., Chuchalin, A., Dolen, W.K., Emuzyte, R., Grouse, L., Humbert, M., Jackson, C., Johnston, S.L., Keith, P.K., Kemp, J.P., Klossek, J.M., Larenas-Linnemann, D., Lipworth, B., Malo, J.L., Marshall, G.D., Naspitz, C., Nekam, K., Niggemann, B., Nizankowska-Mogilnicka, E., Okamoto, Y., Orru, M.P., Potter, P., Price, D., Stoloff, S.W., Vandenplas, O., Viegi, G., Williams, D., 2008. Allergic rhinitis and its impact on asthma (ARIA) 2008. *Allergy* 63, 8–160.
- Cazzoletti, L., Ferrari, M., Olivieri, M., Verlato, G., Antonicelli, L., Bono, R., de Marco, R., 2015. The gender, age and risk factor distribution differs in self-reported allergic and non-allergic rhinitis: a cross-sectional population-based study. *Allergy, Asthma Clinical Immunol.* 11, 36.
- Choi W and Tulu B (2017). Effective Use of User Interface and User Experience in an mHealth Application. Proceedings of the 50th Hawaii International Conference on System Sciences. <http://dx.doi.org/10.24251/HICSS.2017.460>.
- Deloitte (2017). Smart everything, everywhere: Mobile Consumer Survey 2017. <https://www2.deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/deloitte-au-tmt-mobile-consumer-survey-2017-211117-final.pdf>
- De Weger, L.A., Hiemstra, P.S., Op den Buysch, E., van Vliet, A.J., 2014. Spatiotemporal monitoring of allergic rhinitis symptoms in The Netherlands using citizen science. *Allergy* 69 (8), 1085–1091.
- Erbas, B., Akram, M., Dharmage, S.C., Tham, R., Dennekamp, M., Newbiggin, E., Taylor, P., Tang, M.L.K., Abramson, M.J., 2012. The role of seasonal grass pollen on childhood asthma emergency department presentations. *Clin. Exp. Allergy* 42 (5), 799–805.
- Erbas, B., Chang, J.H., Dharmage, S., Ong, E.K., Hyndman, R., Newbiggin, E., Abramson, M., 2007. Do levels of airborne grass pollen influence asthma hospital admissions? *Clin. Exp. Allergy* 37 (11), 1641–1647.
- Haberle, S., Bowman, D., Newnham, R., Johnston, F., Beggs, P., Buters, J., Campbell, B., Erbas, B., Godwin, I., Green, B.J., Huete, A., Jaggard, A.K., Medek, D., Murray, F., Newbiggin, E., Thibaudon, M., Vicendese, D., Williamson, G.J., Davies, J.M., 2014. The macroecology of airborne pollen in Australian and New Zealand urban areas. *PLoS One* 9, e97925.
- Hanigan, I., Johnston, F., 2007. Respiratory hospital admissions were associated with ambient airborne pollen in Darwin, Australia, 2004–2005. *Clin. Exp. Allergy* 37 (10), 1556–1565. <https://doi.org/10.1111/j.1365-2222.2007.02800.x>.
- Huynh, B.T., Tual, S., Turbelin, C., Pelat, C., Cecchi, L., D'Amato, G., Blanchon, T., Annesi-Maesano, I., 2010. Short-term effects of airborne pollens on asthma attacks as seen by general practitioners in the Greater Paris area, 2003–2007. *Prim. Care Respir. J.* 19 (3), 254–259.
- Johnsen, C.R., Weeke, E.R., Nielsen, J., Jensen, J., Mosbech, H., Frølund, L., Madsen, F., Poulsen, L.K., 1992. Aeroallergen analyses and their clinical relevance. 2. Sampling by high-volume air sampler with immunochemical quantification versus Burkard pollen trap sampling with morphologic quantification. *Allergy* 47, 510–516.
- Johnston, F., Hanigan, I., Bowman, D., 2009. Pollen loads and allergic rhinitis in Darwin, Australia: A potential health outcome of the grass-fire cycle. *EcoHealth* 6, 99–108.
- Johnston, F.H., Wheeler, A.J., Williamson, G.J., Campbell, S.L., Jones, P.J., Koolhof, I.S., Lucani, C., Cooling, N.B., Bowman, D.M.J.S., 2018. Using smartphone technology to reduce health impacts from atmospheric environmental hazards. *Environ. Res. Lett.* 13 (4), 044019.
- Karatzas, K., Voukantsis, D., Jaeger, S., Berger, U., Smith, M., Brandt, O., Zuberbier, T., Bergmann, K., 2014. The patient's hay-fever diary: three years of results from Germany. *Aerobiologia* 30 (1), 1–11.
- Kiotseridis, H., Cilio, C.M., Bjermer, L., Tunsäter, A., Jacobsson, H., Dahl, A., 2013. Grass pollen allergy in children and adolescents-symptoms, health related quality of life and the value of pollen prognosis. *Clinic. Trans. Allergy* 3 (1), 19.
- Marks, G.B., Colquhoun, J.R., Girgis, S.T., Hjelmroos Koski, M., Treloar, A.B.A., Hansen, P., Downs, S.H., Car, N.G., 2001. Thunderstorm outflows preceding epidemics of asthma during spring and summer. *Thorax* 56, 468–471.
- Medek, D.E., Kljakovic, M., Fox, I., Pretty, D.G., Prebble, M., 2012. Hay fever in a changing climate: linking an internet-based diary with environmental data. *EcoHealth* 9 (4), 440–447.
- Meltzer, E.O., Bukstein, D.A., 2011. The economic impact of allergic rhinitis and current guidelines for treatment. *Ann. Allergy Asthma Immunol.* 106 (2), S12–S16.
- Nathan, R.A., Meltzer, E.O., Seiner, J.C., Storms, W., 1997. Prevalence of allergic rhinitis in the United States. *J. Allergy Clinical Immunol.* 99 (6), S808–S814.
- Olsson, P., Berglund, N., Bellander, T., Stjärne, P., 2003. Prevalence of self-reported allergic and non-allergic rhinitis symptoms in Stockholm: relation to age, gender, olfactory sense and smoking. *Acta oto-laryngologica* 123 (1), 75–80.
- Ong, E.K., Singh, M.B., Knox, R.B., 1995. Grass pollen in the atmosphere of Melbourne: Seasonal distribution over nine years. *Grana* 34, 58–63.
- R Core Team, 2015. A language and environment for statistical computing.
- Schäppi, G.F., Taylor, P.E., Kenrick, J., Staff, I.A., Suphioglu, C., 1998. Predicting the grass pollen count from meteorological data with regard to estimating the severity of hay fever symptoms in Melbourne Australia. *Aerobiologia* 14, 29–37.
- Silver, J.D., Sutherland, M.F., Johnston, F.H., Lampugnani, E.R., McCarthy, M.A., Jacobs, S.J., Pezza, A.B., Newbiggin, E.J., 2018. Seasonal asthma in Melbourne, Australia, and some observations on the occurrence of thunderstorm asthma and its predictability. *PLoS One* 13, (4) e0194929.

- Steiger, J.H., 1980. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87 (2), 245.
- Tan, R., Cvetkovski, B., Kritikos, V., Price, D., Yan, K., Smith, P., Bosnic-Anticevich, S., 2018. The burden of rhinitis and the impact of medication management within the community pharmacy setting. *J. Allergy Clin. Immunol. Pract.* 6 (5), 1717–1725.
- Thien, F., Beggs, P.J., Csutoros, D., Darvall, J., Hew, M., Davies, J.M., Bardin, P.G., Bannister, T., Barnes, S., Bellomo, R., Byrne, T., Casamento, A., Conron, M., Cross, A., Crosswell, A., Douglass, J.A., Durie, M., Dyett, J., Ebert, E., Erbas, B., French, C., Gelbart, B., Gillman, A., Harun, N.S., Huete, A., Irving, L., Karalapillai, D., Ku, D., Lachapelle, P., Langton, D., Lee, J., Looker, C., MacIsaac, C., McCaffrey, J., McDonald, C.F., McGain, F., Newbiggin, E., O'Hehir, R., Pilcher, D., Prasad, S., Rangamuwa, K., Ruane, L., Sarode, V., Silver, J.D., Southcott, A.M., Subramaniam, A., Suphioglu, C., Susanto, N.H., Sutherland, M.F., Taori, G., Taylor, P., Torre, P., Vetro, J., Wigmore, G., Young, A.C., Guest, C., 2018. The Melbourne epidemic thunderstorm asthma event 2016: an investigation of environmental triggers, effect on health services, and patient risk factors. *Lancet Planet. Health* 2, E255–E263.
- Walter, L.M., Tamanyan, K., Nisbet, L., Weichard, A.J., Davey, M.J., Nixon, G.M., Horne, R.S.C., 2019. Pollen levels on the day of polysomnography influence sleep disordered breathing severity in children with allergic rhinitis. *Sleep Breath* 23, 651–657.
- Wang, D.Y., 2005. Risk factors of allergic rhinitis: genetic or environmental? *Ther. Clin. Risk Manag.* 1 (2), 115–123.
- Werchan, B., Werchan, M., Mücke, H.-G., Bergmann, K.C., 2018. Spatial distribution of pollen-induced symptoms within a large metropolitan area—Berlin, Germany. *Aerobiologia* 34, 539–556.
- WHO (2012). Management of patient information: trends and challenges in Member States: based on the findings of the second global survey on eHealth. World Health Organization. URL: http://apps.who.int/iris/bitstream/10665/76794/1/9789241504645_eng.pdf
- Williams, E.J., 1959. The comparison of regression variables. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* 21 (2), 396–399.

A

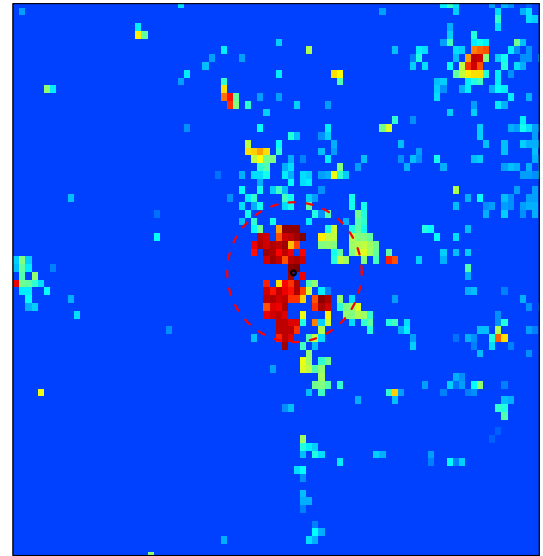
Melbourne

Population density



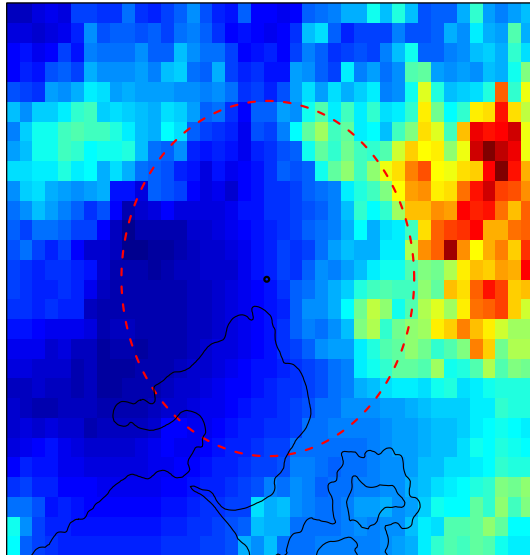
B

Canberra

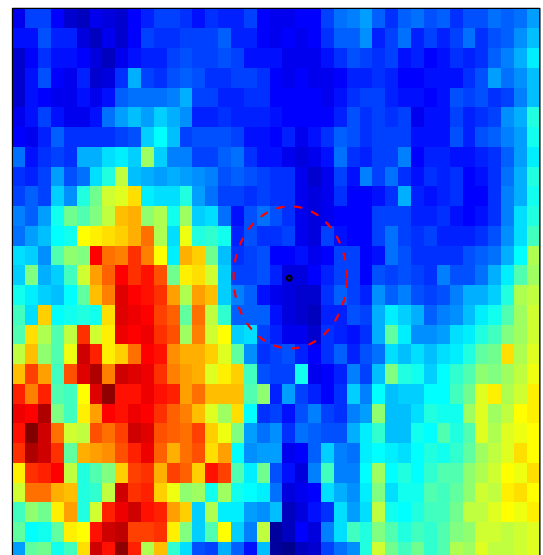


C

Rainfall

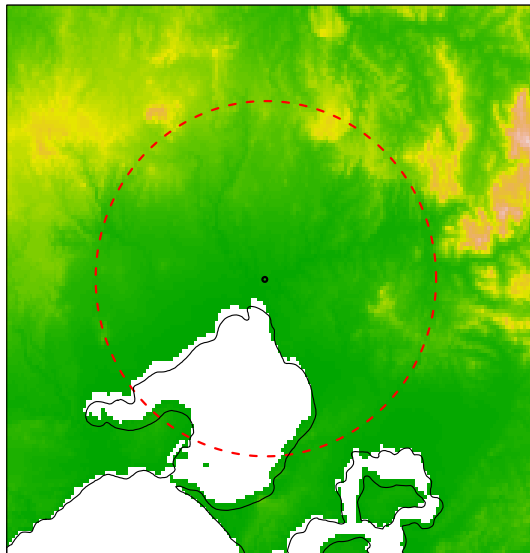


D

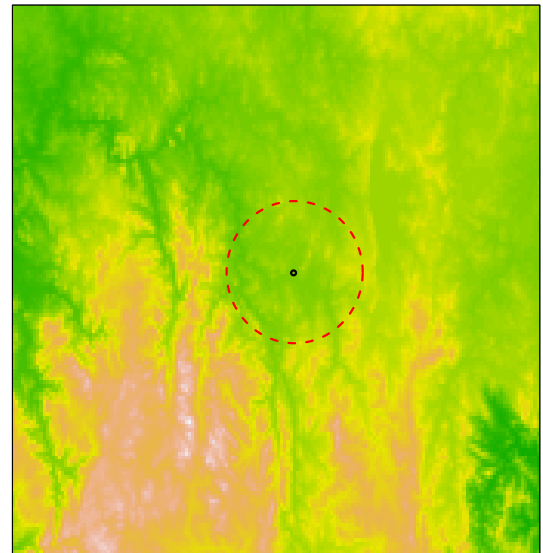


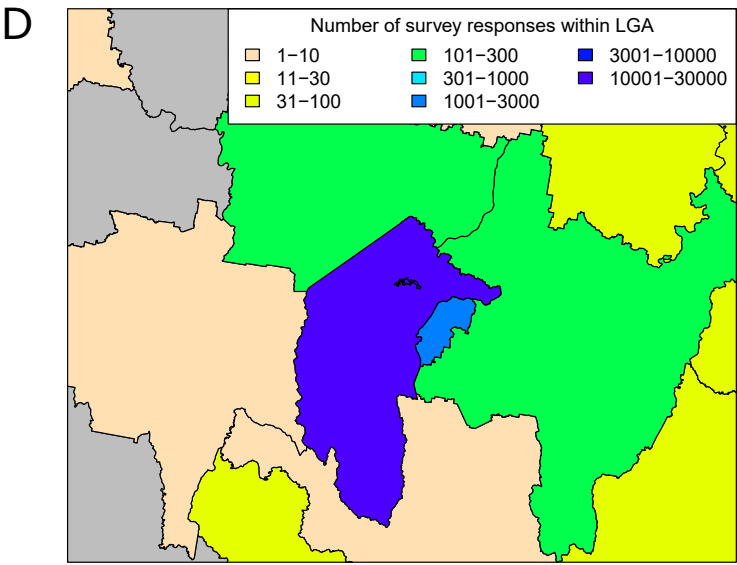
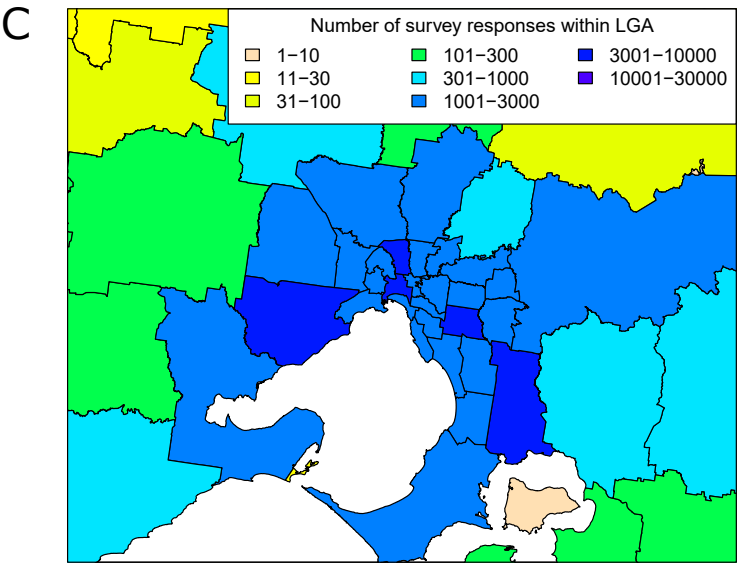
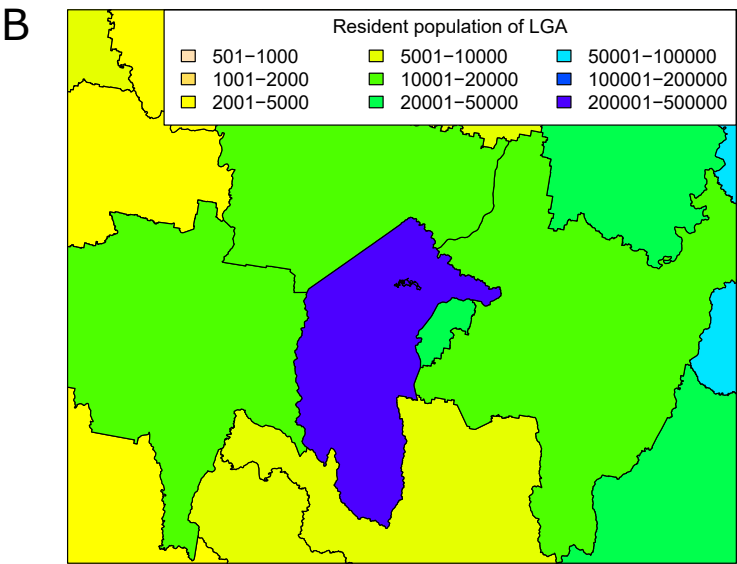
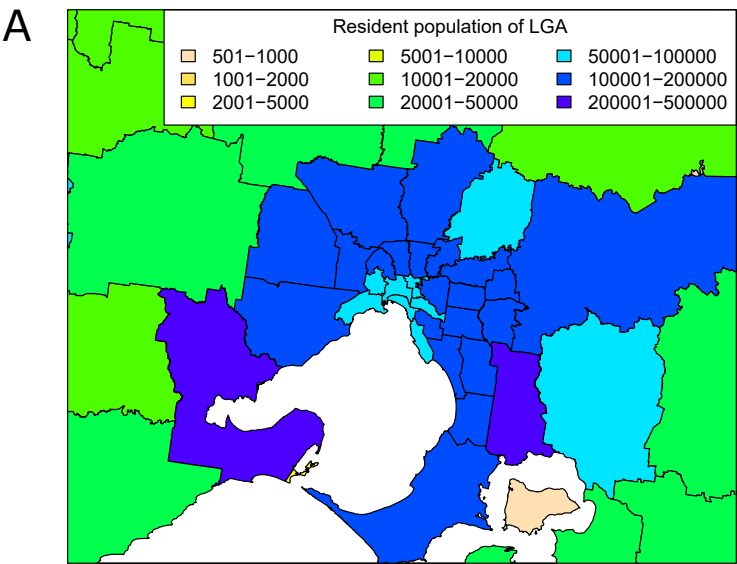
E

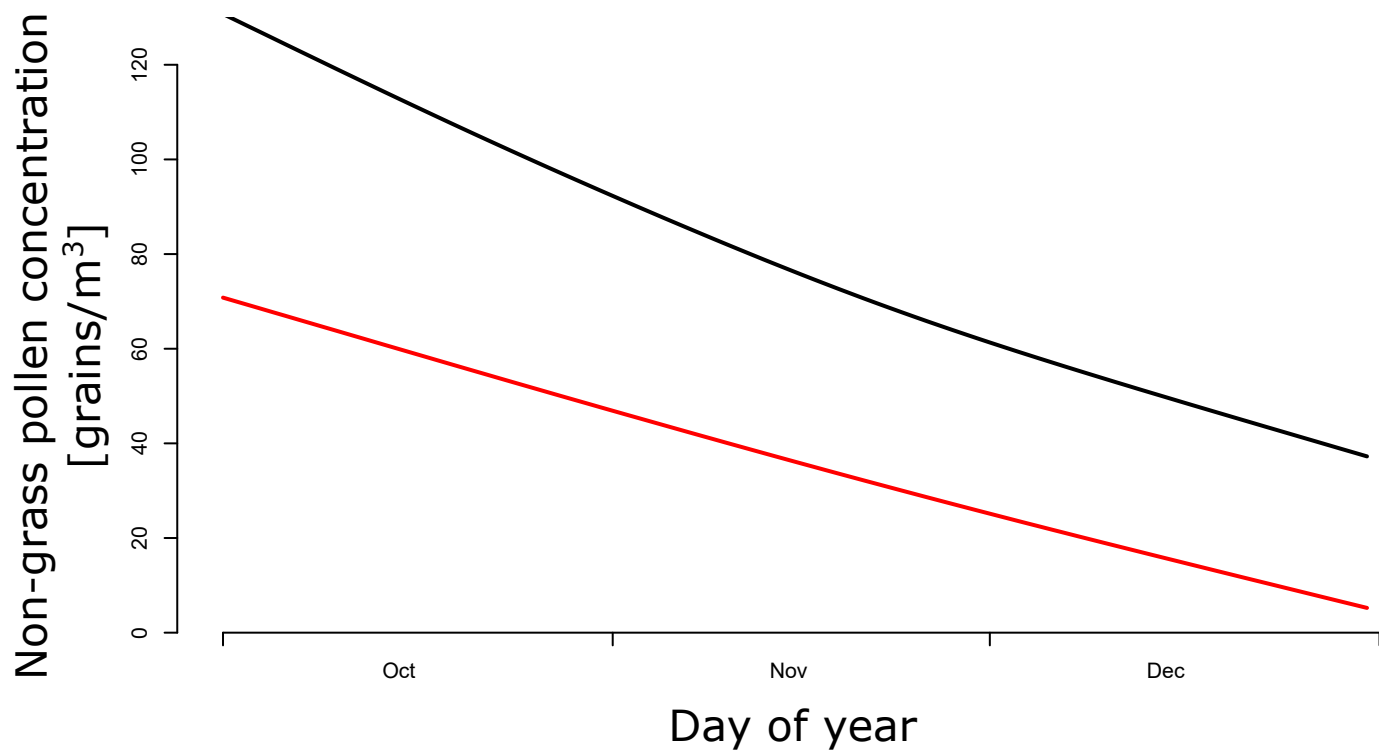
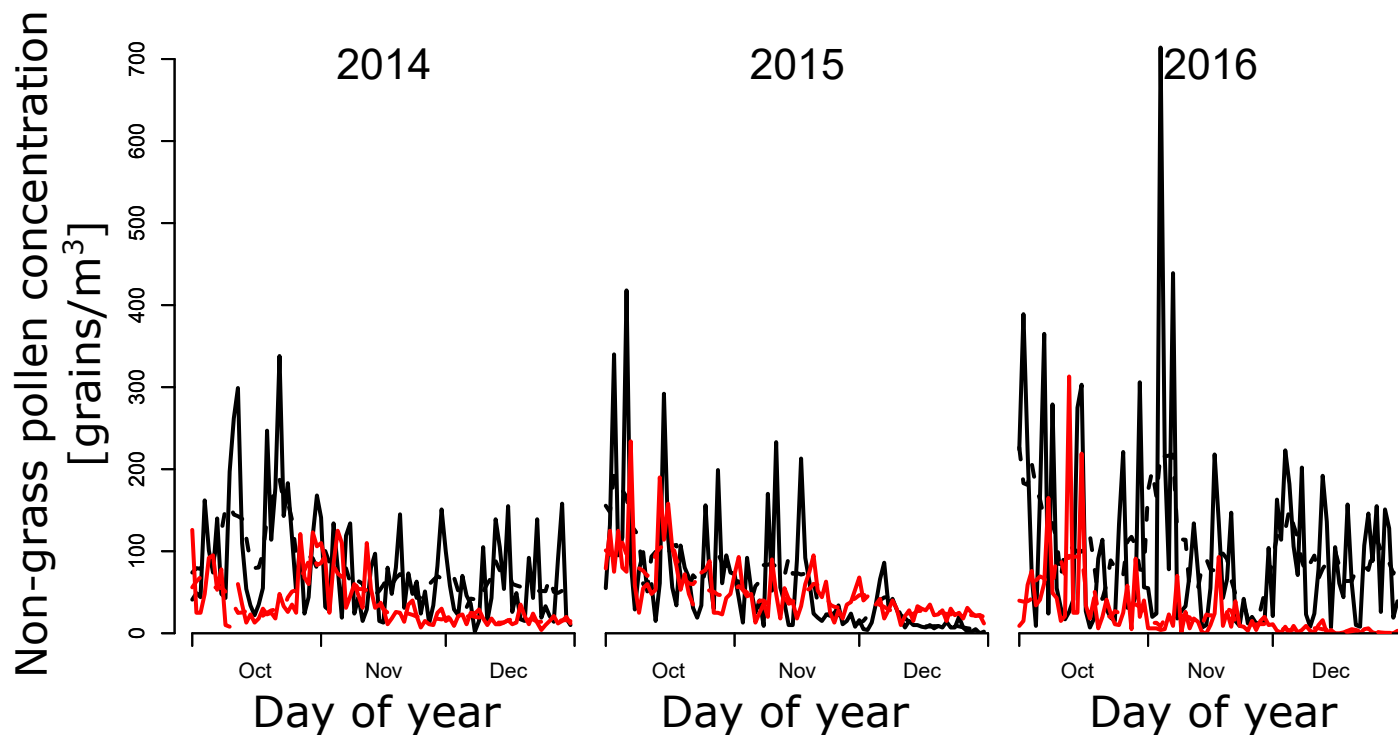
Terrain



F

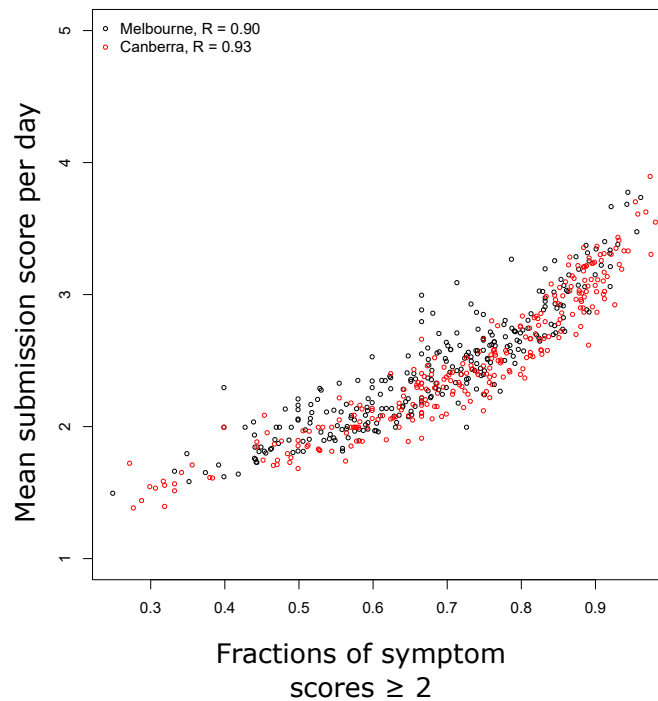




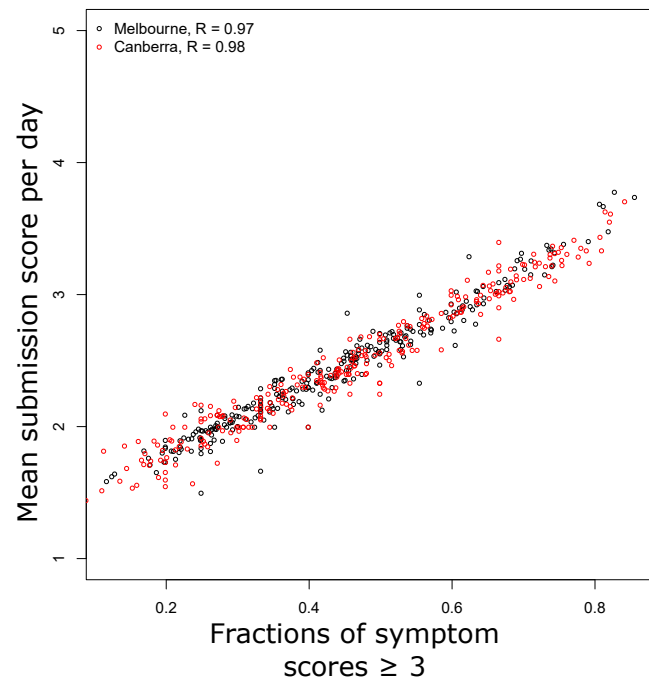


■ Melbourne ■ Canberra — Original - - Moving average

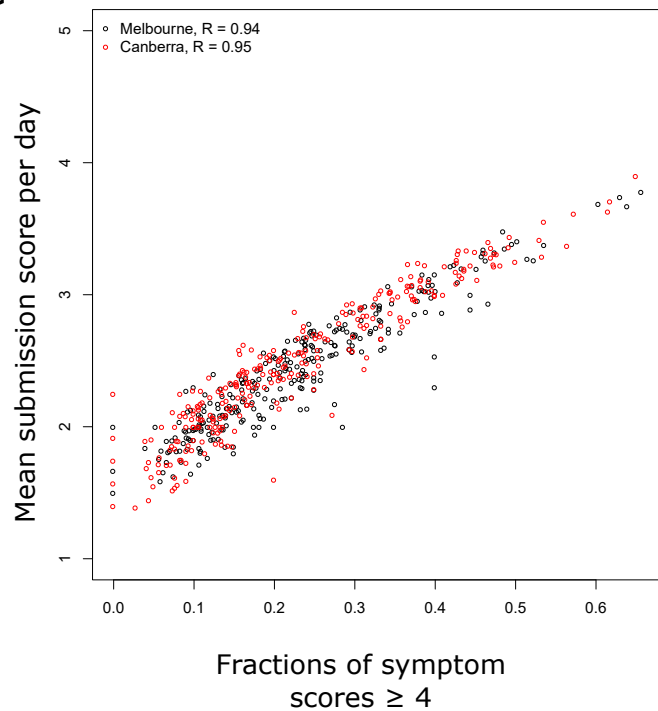
A



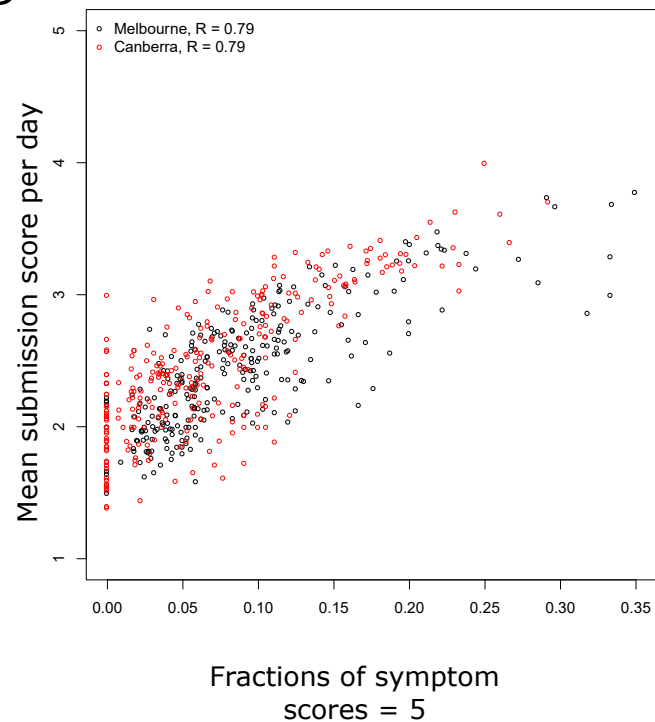
B



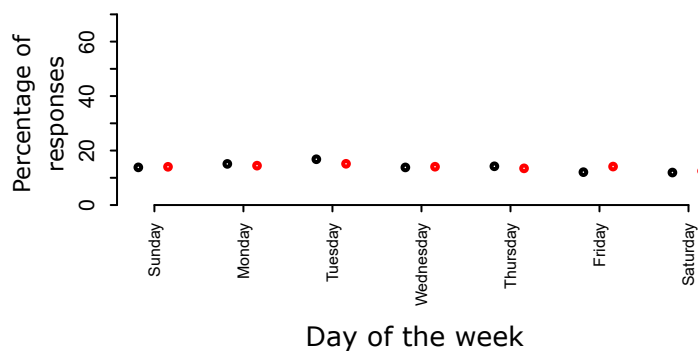
C



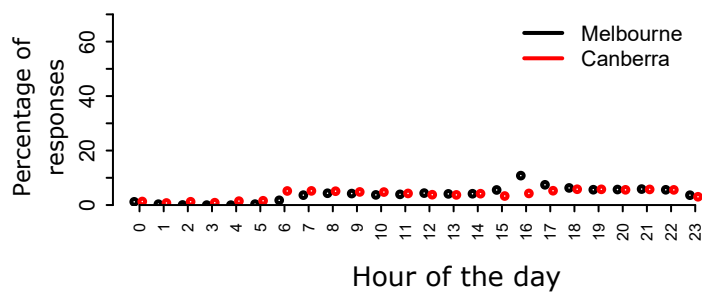
D



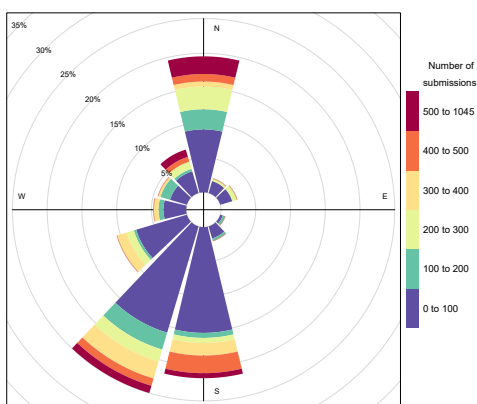
A



B

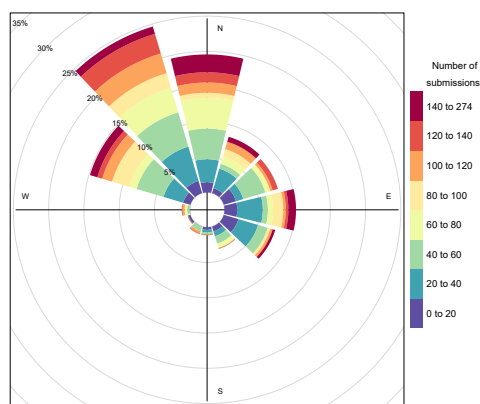


C



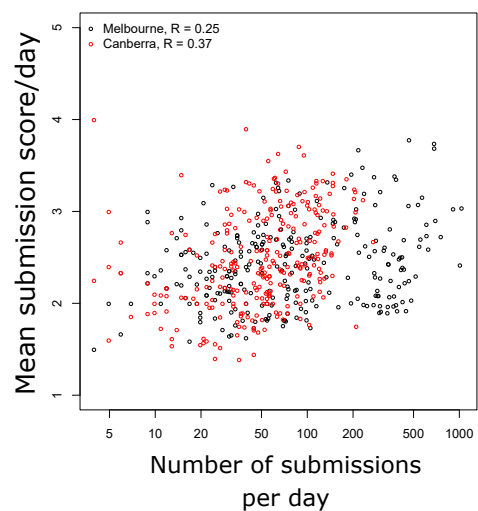
Number of submissions
(Melbourne)

D

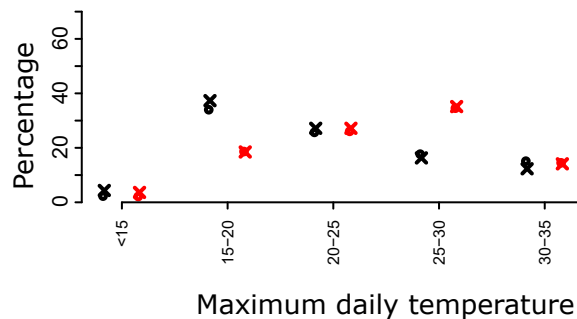


Number of submissions
(Canberra)

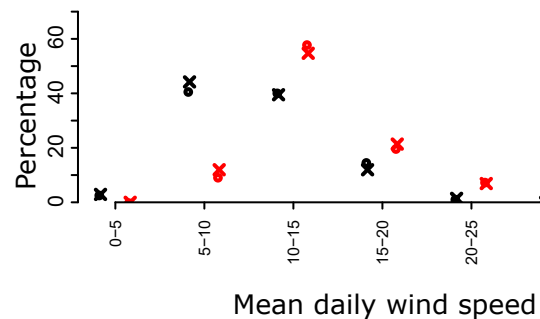
E



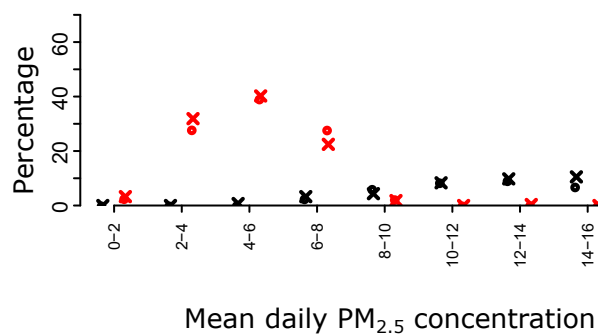
A



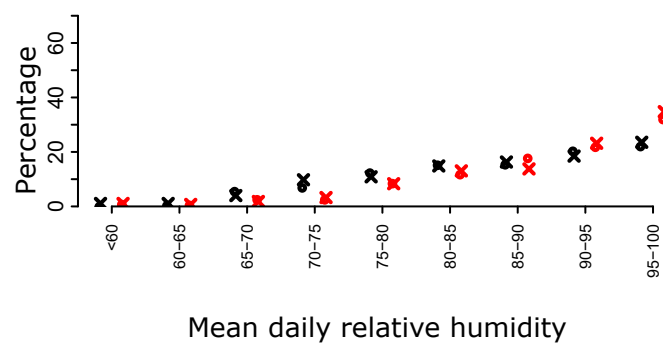
B



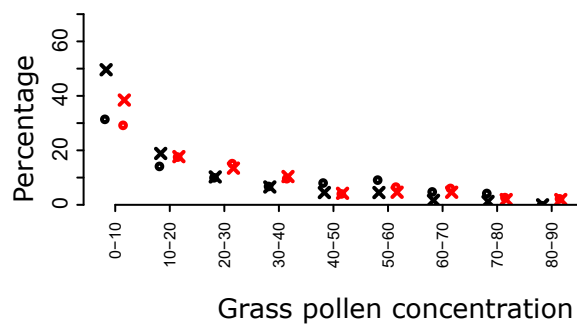
C



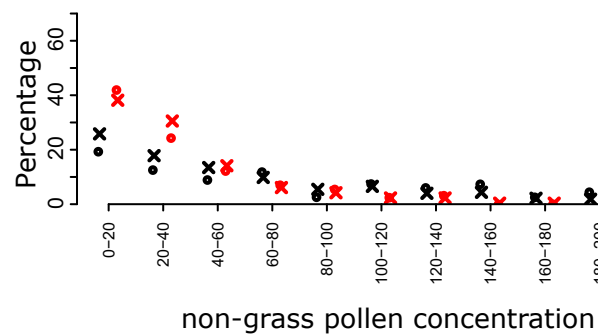
D



E

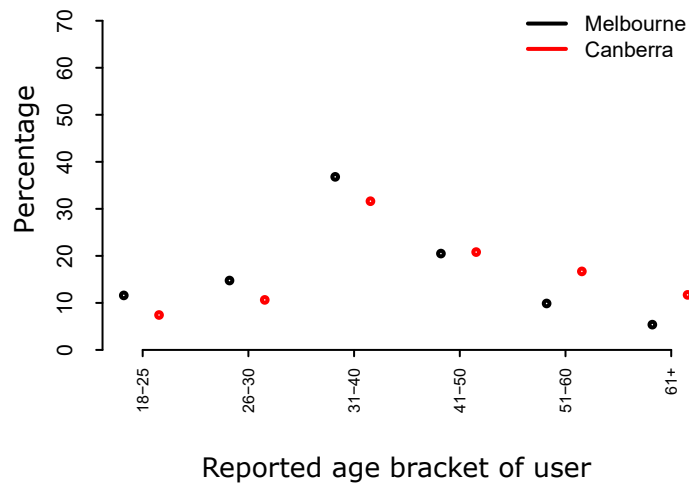


F

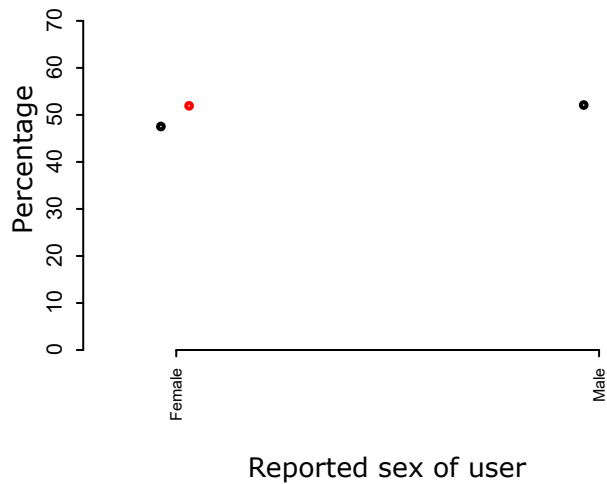


— Melbourne • Percentage of responses
 — Canberra x Percentage of days

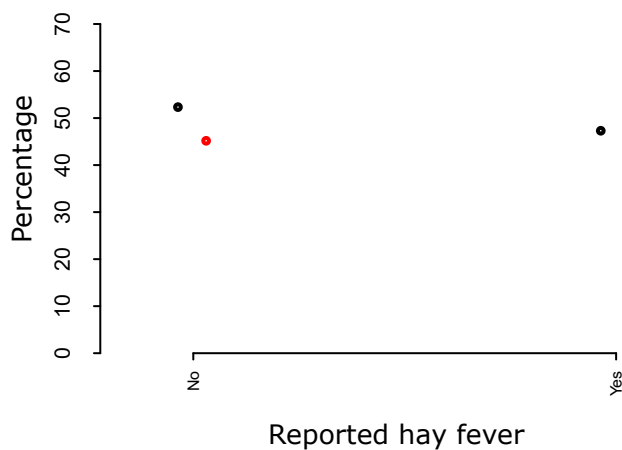
A



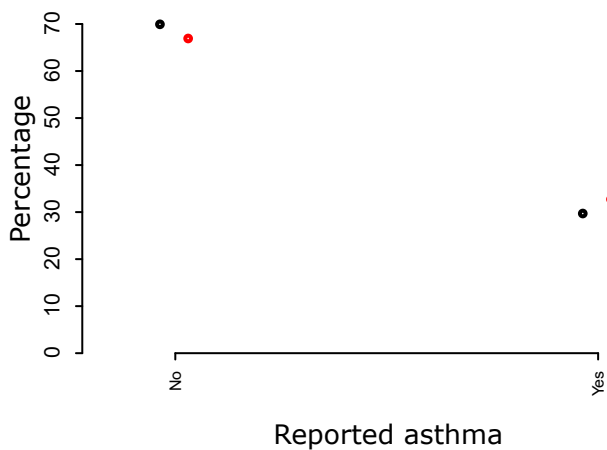
B



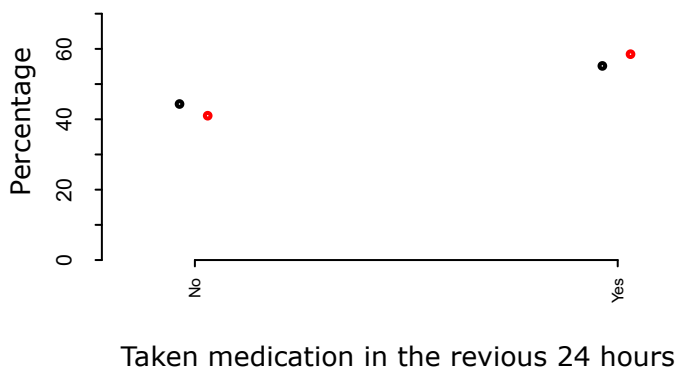
C



D



E



F

