The association between air pollution and the incidence of idiopathic pulmonary fibrosis in Northern Italy

Sara Conti, Sergio Harari, Antonella Caminati, Antonella Zanobetti, Joel D Schwartz, Pietro A Bertazzi, Giancarlo Cesana, Fabiana Madotto

SUPPLEMENTARY MATERIAL

Supplement to the methods section

Assessment of the PM_{10} municipality-specific concentrations

We assessed which cells of the grid partly or totally overlaid each municipality, and computed the municipality-specific average daily PM_{10} concentration as follows:

$$\bar{y}_{il} = \frac{\sum_{j=1}^{k} w_j \bar{y}_{jl}}{\sum_{j=1}^{k} w_j},$$

Where $\bar{y}_{il}$ is the daily average concentration of PM_{10} in the i-th municipality during the *l*-th day, $\bar{y}_j$ is the daily average concentration estimated in the *j*-th cell that overlays the *i*-th municipality during the *l*-th day, *k* is the number of cells that overlay the municipality, and

$$w_j = \frac{overlay\ area\ i-th\ municipality\ and\ j-th\ cell}{area\ i-th\ municipality} * 100.$$

Where "overlay area *i*-th municipality and *j*-th cell" is the area (in Km^2) of the *i*-th muncipality that lies within the *j*-th cell, and "area *i*-th municipality" is the total area (in km^2) of the *i*-th municipality.

The thus obtained daily concentrations were then averaged for each municipality over different time windows (2000-2009, 2001-2009, 2002-2009, 2003-2009, 2004-2009, 2005-2009) .

Assessment of the completeness of NO_2 and O_3 monitoring stations

In order to asses which NO_2 monitors to select in order to compute the municipality-specific concentrations, we analyzed the completeness of each monitor following the subsequent procedure:

- if during a day at least 75% of the hourly measures (18 hours) were available, we considered that day complete and calculated the daily average NO_2 concentration;
- we computed the number of complete days for each month and we considered complete a month with at least 75% of the measures;
- we computed the number of complete months for each year and we considered complete one year with at least 10 complete months.

Similarly to the procedure we adopted for NO_2, we analyzed the completeness of O_3 monitors following the subsequent procedure:

- if during a day at least 75% of the hourly measures (18 hours) were available, we considered that day complete and calculated the daily 8-hours maximum O_3 concentration;
- we computed the number of complete days for each month and we considered complete a month with at least 75% of the measures;
- we focused on the warm season and we computed the number of complete months within that season for each year, we then considered complete one year with at least 5 complete months.

Finally, for both pollutants, for each municipality we averaged the concentration over different time windows (2000-2010, 2001-2010, 2002-2010, 2003-2010, 2004-2010, 2005-2010), each time including only monitors with time-series that could be considered complete for at least 75% of the years.

Selection of the time-windows for estimating the chronic exposure to air pollutants

Finally, we computed the percentage of cases and municipalities for which we could estimate NO_2 and O_3 concentrations, since some municipalities are located at a distance greater than 10 km from the selected stations, and we averaged exposures over different time windows. Given that coverage grew with time, while average exposures remained fairly stable over the time windows, we decided to use the average concentrations over 2005-2010 for NO_2 and O_3, and over 2005-2009 for PM_{10} as proxies for chronic exposures (see supplementary material).

Details on negative binomial models

During our analysis, for each pollutant, we first built an unadjusted model:

$$\log E(Y_i) = \log N_i + \beta_0 + \beta_1 pollutant_i$$

where, within the i-th aggregated area, Y_i is the cumulative count of incident cases over the study period and N_i is the person-years at risk estimated on the 2005–2010 average population; $pollutant_i$ is the average of the pollutant examined.

We then built a fully adjusted model, accounting for all available confounders:

$$\log E(Y_i) = \log N_i + \beta_0 + \beta_1\, pollutant_i + \beta_2\, p_{males,i} + \beta_3\, age_i + \beta_4\, income_i + \beta_5\, I(agricultural)_i + \beta_6\, I(urban/industrial)_i + \beta_7\, temperature_i$$

where, within the *i*-th aggregated area, $p_{males,i}$, age_i and $income_i$ are respectively the average proportion of males, the mean age and the average per-capita income; $I(agricultural)_i$ and $I(urban/industrial)_i$ are indicator variables for the main use of land (with natural land as the reference category); $temperature_i$ is the average temperature over the study period.

Finally, we also built the best model, that includes only confounders that improved the model fitting, that we evaluated based on the Akaike Information Criterion (AIC): all confounders that lowered the AIC where retained in the final model.

Sensitivity analysis - Alternative case definitions

The generic case definition is less stringent as compared to the broad case definition (BCD): it requires that the subject has at least one hospital admission or outpatient visit with a diagnosis of IPF during the study period.

The narrow case definition is more stringent than the BCD: in addition to satisfy BCD criteria, it also requires the subject to have at least one surgical lung biopsy (ICD-9-CM code 33.28), transbronchial lung biopsy (ICD-9-CM codes 33.27) or computed tomography of the thorax (ICD-9-CM code 87.41) before the last traceable IPF diagnosis.

Supplement to the results section

Evaluation of the distribution of the outcome and confounders at a municipality level.

Table S1. Distribution of the outcome and of the potential confounders in the municipalities of Lombardy.

	Descriptive statistics
Number of municipalities	1545
Number of IPF cases	2093
N° IPF cases per municipality	
Mean (SD*)	1.4 (11.34)
Median (IQR†)	0 (0; 1)
Minimum; Maximum	0; 433
CV‡	8.37
Average population per municipality	
Mean (SD*)	6,249 (34,855)
Median (IQR†)	2,609 (1,152; 5,666)
Minimum; Maximum	36; 1,311,775
CV‡	5.58
Percentage of males per municipality	
Mean (SD*)	49.6 (1.4)
Median (IQR†)	49.5 (48.8; 50.3)
Minimum; Maximum	42.9; 58.8
CV‡	0.03
Mean age per municipality	
Mean (SD*)	42.6 (3.1)
Median (IQR†)	42.2 (40.6; 44.2)
Minimum; Maximum	32.4; 59.0
CV‡	0.07
Average income (in 1000 euros) per municipality	
Mean (SD*)	18.7 (3.3)
Median (IQR†)	18.6 (16.6; 20.5)
Minimum; Maximum	6.5 - 46.6
CV‡	0.17
Main use of the land within the municipalilty - N(%)	
Natural	452 (29.26)
Agricultural	680 (44.01)
Urban and industrial	413 (26.73)
Average temperature (°C)† per municipality	
Mean (SD*)	12 (2.4)
Median (IQR†)	13 (11.1; 13.2)
Minimum; Maximum	0; 14
CV‡	0.20

* SD: Standard Deviation
† Interquartile Range
‡ Coefficient of Variation

Evaluation of the coverage and average exposure to PM_{10}, NO_2 and O_3 over different time windows.

Table S2. Descriptive statistics of the municipality-specific average of the daily PM_{10} concentration, over different time windows.

	Average PM_{10} concentration ($\mu g/m^3$)		
	Overall	**Warm season**	**Cold season**
2000 - 2009			
Municipalities with missing exposure - N(%)	14 (0.91%)	14 (0.91%)	14 (0.91%)
Mean (SD)	41 (3.87)	29 (2.93)	53 (5.10)
Median (IQR)	42 (38.4 - 43.8)	30 (27.5 - 31.5)	53 (49.3 - 57.2)*
Min - Max	32 - 51	22 - 37	41 - 66
2001 - 2009			
Municipalities with missing exposure - N(%)	14 (0.91%)	14 (0.91%)	14 (0.91%)
Mean (SD)	41 (4.00)	29 (2.96)	54 (5.30)
Median (IQR)	42 (38.3 - 44.0)	30 (26.8 - 31.1)	54 (49.4 - 57.7)*
Min - Max	32 - 51	21 - 36	41 – 67
2002 - 2009			
Municipalities with missing exposure - N(%)	14 (0.91%)	14 (0.91%)	14 (0.91%)
Mean (SD)	41 (4.11)	29 (2.98)	53 (5.47)
Median (IQR)	42 (37.9 - 43.9)	30 (26.7 - 31.2)	54 (48.6 - 57.4)*
Min - Max	31 - 51	21 - 36	40 - 66
2003 - 2009			
Municipalities with missing exposure - N(%)	14 (0.91%)	14 (0.91%)	14 (0.91%)
Mean (SD)	41 (4.09)	29 (2.99)	53 (5.44)
Median (IQR)	41 (37.3 - 43.6)	29 (26.6 - 31.2)	53 (47.7 - 56.7)*
Min - Max	32 - 51	22 - 35	40 - 66
2004 - 2009			
Municipalities with missing exposure - N(%)	14 (0.91%)	14 (0.91%)	14 (0.91%)
Mean (SD)	40 (4.18)	28 (3.00)	52 (5.55)
Median (IQR)	41 (36.5 - 42.8)	28 (25.5 - 30.0)	53 (47.1 - 56.2)*
Min - Max	31 - 50	21 - 34	40 - 65
2005 - 2009			
Municipalities with missing exposure - N(%)	14 (0.91%)	14 (0.91%)	14 (0.91%)
Mean (SD)	39 (4.31)	27 (3.07)	52 (5.76)
Median (IQR)	40 (35.8 - 42.4)	28 (24.7 - 29.5)	53 (46.7 - 56.0)*
Min - Max	30 - 50	20 - 34	40 - 66

*p-value Wilcoxon rank-signed test vs "Warm season" <0.05
Abbreviations: SD: Standard Deviation; IQR: Interquartile Range

Table S3. Descriptive statistics of the municipality-specific average of the daily NO_2 concentration over different time windows.

	Average NO_2 concentration ($\mu g/m^3$)		
	Overall	**Warm season**	**Cold season**
2000 - 2010			
Municipalities with missing exposure - N(%)	705 (45.63%)	705 (45.63%)	705 (45.63%)
Mean (SD)	40 (10.14)	30 (9.45)	50 (11.10)
Median (IQR)	39 (31.3 - 48.7)	29 (21.5 - 37.4)	49 (41.9 - 59.5)*
Min - Max	16 - 67	9 - 59	23 - 75
2001 - 2010			
Municipalities with missing exposure - N(%)	656 (42.46%)	656 (42.46%)	656 (42.46%)
Mean (SD)	40 (9.76)	29 (9.05)	50 (10.77)
Median (IQR)	39 (30.9 - 48.1)	28 (21.5 - 36.8)	49 (41.5 - 59.4)*
Min - Max	16 - 67	9 - 58	23 - 75
2002 - 2010			
Municipalities with missing exposure - N(%)	646 (41.81%)	646 (41.81%)	646 (41.81%)
Mean (SD)	39 (9.43)	29 (8.73)	50 (10.47)
Median (IQR)	39 (31.1 - 47.3)	28 (21.9 - 36.1)	48 (41.3 - 58.7)*
Min - Max	16 - 66	9 - 57	23 - 75
2003 - 2010			
Municipalities with missing exposure - N(%)	587 (37.99%)	587 (37.99%)	587 (37.99%)
Mean (SD)	39 (9.42)	28 (8.74)	49 (10.48)
Median (IQR)	39 (31.2 - 46.7)	28 (22.1 - 34.5)	49 (41.3 - 57.4)*
Min - Max	16 - 66	9 - 57	23 - 75
2004 - 2010			
Municipalities with missing exposure - N(%)	541 (35.02%)	541 (35.02%)	541 (35.02%)
Mean (SD)	38 (9.22)	27 (8.18)	49 (10.51)
Median (IQR)	38 (30.3 - 45.3)	27 (21.2 - 32.3)	48 (41.9 - 56.4)*
Min - Max	16 - 65	9 - 56	24 - 74
2005 - 2010			
Municipalities with missing exposure - N(%)	483 (31.26%)	483 (31.26%)	483 (31.26%)
Mean (SD)	37 (9.03)	26 (7.95)	48 (10.34)
Median (IQR)	38 (30.0 - 43.4)	25 (20.7 - 31.3)	48 (41.1 - 55.3)*
Min - Max	16 - 65	9 - 55	24 - 74

*p-value Wilcoxon rank-signed test vs "Warm season" <0.05
Abbreviations: SD: Standard Deviation; IQR: Interquartile Range

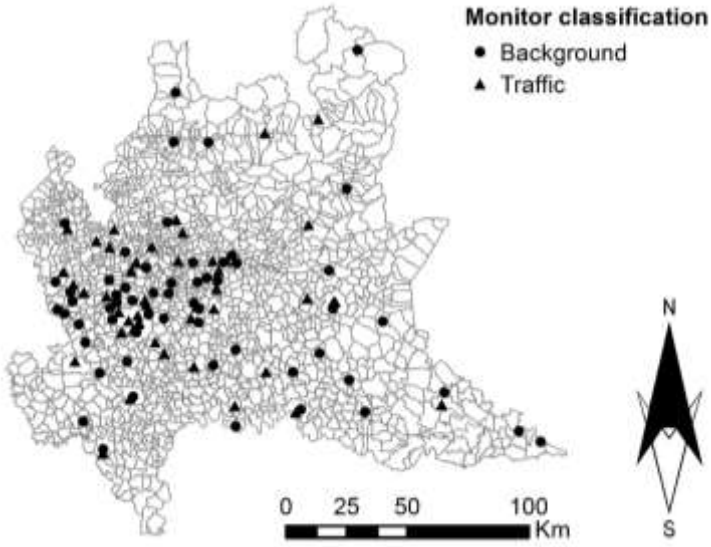
Table S4. Descriptive statistics of the municipality-specific average of the 8-hours maximum O_3 concentration during the warm season, over different time windows.

	Average O_3 concentration (ppm)
2000 - 2010	
Municipalities with missing exposure - N(%)	1,017 (65.83%)
Mean (SD)	112 (7.14)
Median (IQR)	113 (105.1 - 117.8)
Min - Max	94 - 126
2001 - 2010	
Municipalities with missing exposure - N(%)	988 (63.95%)
Mean (SD)	112 (7.41)
Median (IQR)	111 (104.0 - 118.2)
Min - Max	94 - 126
2002 - 2010	
Municipalities with missing exposure - N(%)	931 (60.26%)
Mean (SD)	111 (7.40)
Median (IQR)	112 (103.9 - 118.2)
Min - Max	95 - 126
2003 - 2010	
Municipalities with missing exposure - N(%)	912 (59.03%)
Mean (SD)	111 (7.47)
Median (IQR)	110 (104.4 - 118.7)
Min - Max	95 - 126
2004 - 2010	
Municipalities with missing exposure - N(%)	842 (54.50%)
Mean (SD)	109 (6.69)
Median (IQR)	109 (102.5 - 114.1)
Min - Max	91 - 122
2005 - 2010	
Municipalities with missing exposure - N(%)	654 (42.33%)
Mean (SD)	109 (7.15)
Median (IQR)	110 (103.6 - 116.4)
Min - Max	91 - 120

Abbreviations: SD: Standard Deviation; IQR: Interquartile Range

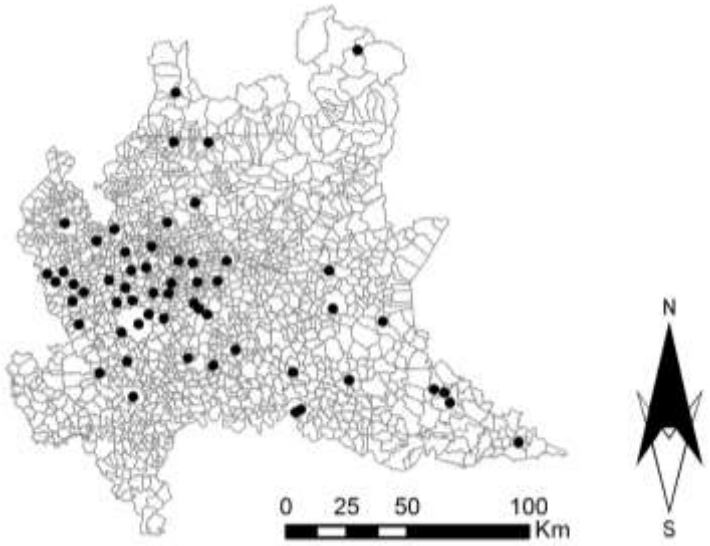
Location of the monitoring stations selected for the analysis

Figure S1. Location of NO_2 background and traffic monitoring stations involved in the analysis. Grey lines delimit municipalities, black dots and triangles represent monitoring stations.



Shapefile of the Italian administrative regions and municipalities downloaded from the website of the Italian national institute of statistics (http://www.istat.it/it/archivio/24613).

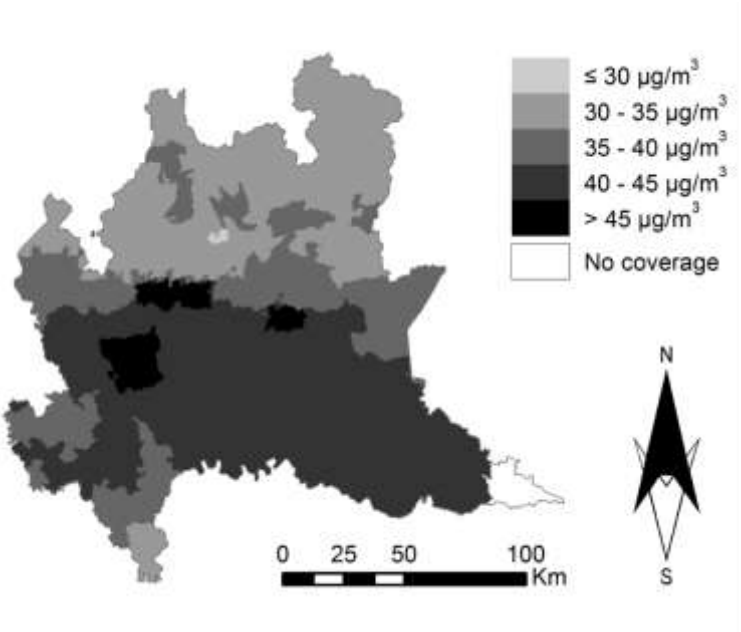
Figure S2. Location of O_3 background monitoring stations involved in the analysis. Grey lines delimit municipalities, black dots represent monitoring stations.



Shapefile of the Italian administrative regions and municipalities downloaded from the website of the Italian national institute of statistics (http://www.istat.it/it/archivio/24613).

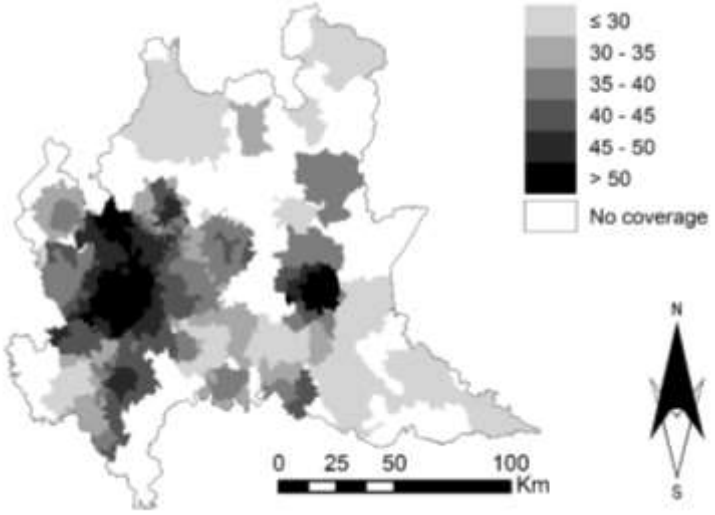
Estimated PM_{10}, NO_2 and O_3 concentrations at a municipality level

Figure S3. Map of the estimated daily average PM_{10} concentration (in $\mu g/m^3$) for the period 2005-2009 at a municipality level.



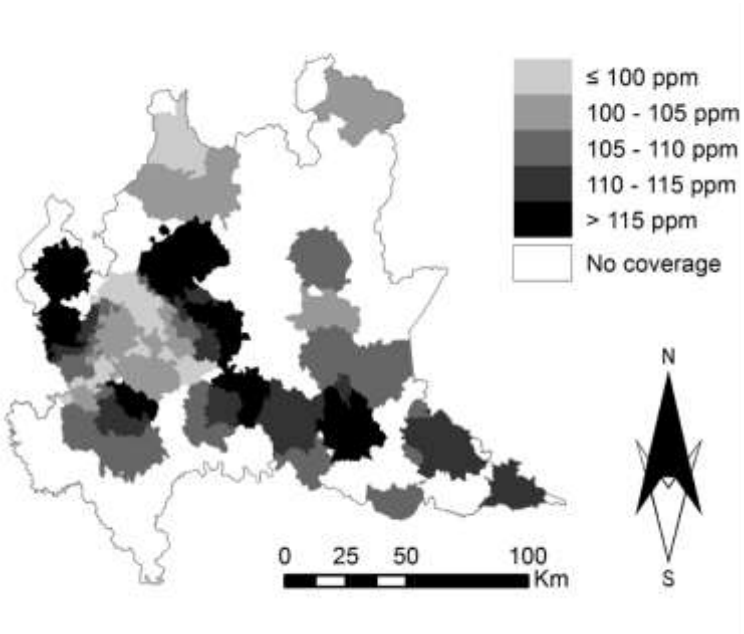
Shapefile of the Italian administrative regions and municipalities downloaded from the website of the Italian national institute of statistics (http://www.istat.it/it/archivio/24613).

Figure S4. Map of the estimated daily average NO_2 concentration (in $\mu g/m^3$) for the period 2005-2010 at a municipality level.



Shapefile of the Italian administrative regions and municipalities downloaded from the website of the Italian national institute of statistics (http://www.istat.it/it/archivio/24613).

Figure S5. Map of the estimated daily average O_3 concentration (in ppm) for the period 2005-2010 at a municipality level.



Shapefile of the Italian administrative regions and municipalities downloaded from the website of the Italian national institute of statistics (http://www.istat.it/it/archivio/24613).

Distribution of the outcome before and after the aggregation.

Figure S6. Distribution of the cumulative count of IPF incident cases at a municipality-level (before the aggregation) and after the aggregation based on PM_{10} concentration.

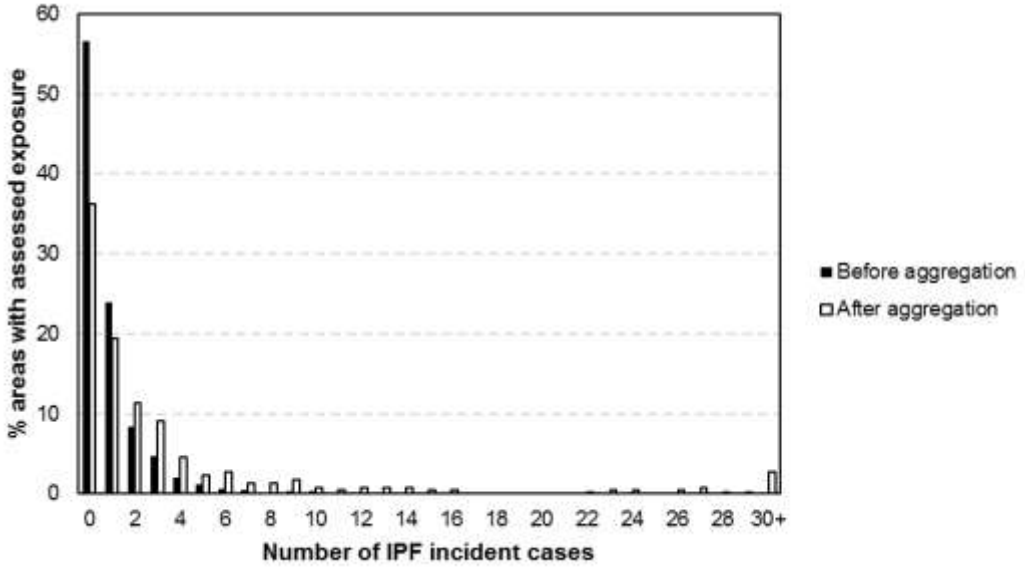


Figure S7. Distribution of the cumulative count of IPF incident cases at a municipality-level (before the aggregation) and after the aggregation based on O_3 concentration.

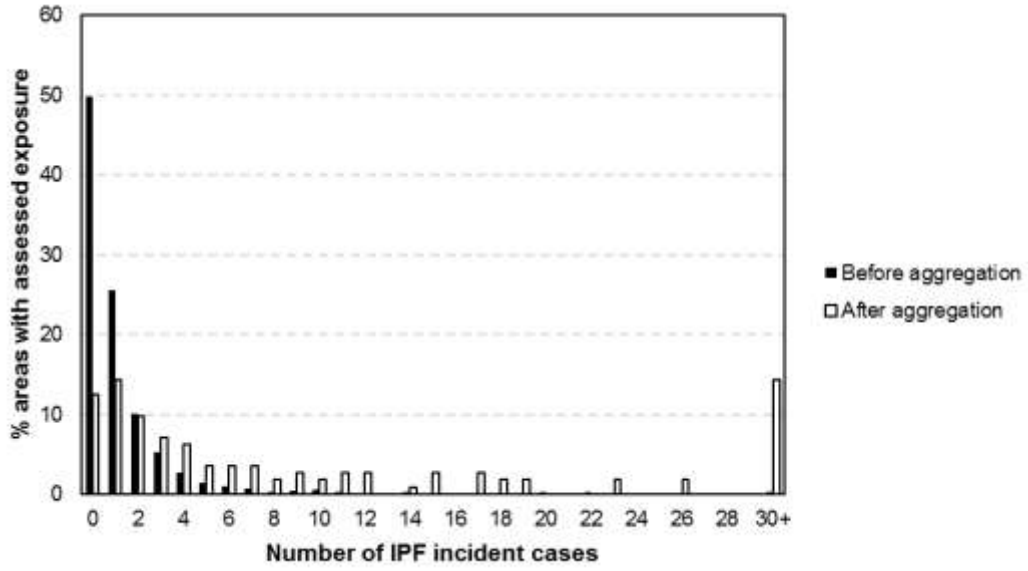
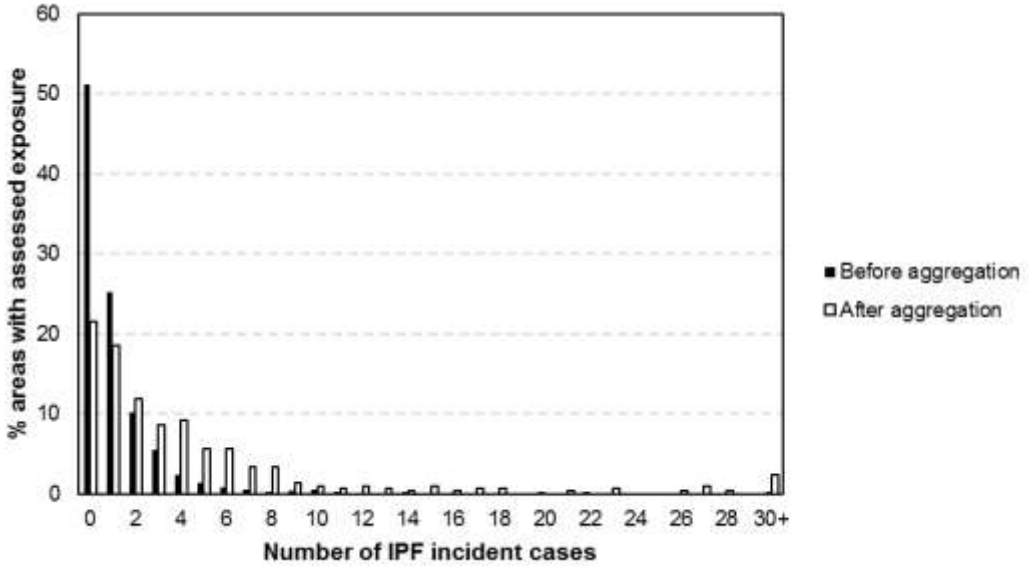


Figure S8. Distribution of the cumulative count of IPF incident cases at a municipality-level (before the aggregation) and after the aggregation based on NO_2 concentration.



Sensitivity analysis

Table S5. Distribution of the outcome in the aggregated areas, according to generic and broad case definition.

	Aggregation based on PM_{10}	**Agggregation based on NO_2**	**Aggregation based on O_3**
Number of areas	398	303	112
N of IPF cases with assessed exposure	2090	1842	1702
Generic Case Definition	2948	2603	2392
Narrow Case Definition	1308	1155	1074
N° IPF cases per area			
Generic Case Definition			
Mean (SD*)	7 (34)	9 (36)	21 (60)
Median (IQR†)	2 (0; 5)	3 (1; 8)	7 (2; 23)
Min; Max	0; 627	0; 608	0; 602
CV‡	4.60	4.19	2.79
Narrow Case Definition			
Mean (SD*)	3 (15)	4 (16)	10 (26)
Median (IQR†)	1 (0; 2)	1 (0; 3)	3 (1; 9)
Min; Max	0; 275	0; 269	0; 265
CV‡	4.57	4.21	2.76

* SD: Standard Deviation
† Interquartile Range
‡ Coefficient of Variation

Table S6. Estimated % change in the Incidence Rate (IR), with related 95% Confidence Interval (CI), for a 10 units increase in the average pollutants concentrations, using generic and narrow case definition.

	Average PM_{10} concentration (10 µg/m^3 increase)	Average O_3 concentration (10 ppm increase)	Average NO_2 concentration (10 µg/m^3 increase)
Generic Case Definition			
Overall			
Unadjusted model	-9.55 (-22.56; 5.66)		8.31 (0.74; 16.45)‡
Fully adjusted model*	-6.24 (-26.45; 19.54)		6.94 (-2.42; 17.20)
Best model†	2.18 (-17.22; 26.12)		7.64 (-0.47; 16.42)§
Cold season			
Unadjusted model	-6.51 (-16.59; 4.80)		8.20 (1.56; 15.29)‡
Fully adjusted model*	-4.92 (-20.70; 14.00)		8.04 (-0.24; 17.01)§
Best model†	1.53 (-12.96; 18.43)		8.26 (0.96; 16.08)‡
Warm season			
Unadjusted model	-14.58 (-31.70; 6.83)	-11.48 (-22.24; 0.77)§	7.71 (-0.72; 16.86)§
Fully adjusted model*	-6.36 (-31.86; 28.70)	-7.04 (-17.91; 5.26)	4.63 (-5.44; 15.78)
Best model†	2.85 (-23.14; 37.64)	-8.40 (-19.07; 3.69)	6.18 (-2.57; 15.70)
Narrow Case Definition			
Overall			
Unadjusted model	-9.37 (-25.87; 10.81)		8.18 (-1.52; 18.83)
Fully adjusted model*	-6.16 (-31.20; 28.00)		5.82 (-5.95; 19.06)
Best model†	1.65 (-21.74; 32.01)		6.19 (-4.09; 17.57)
Cold season			
Unadjusted model	-8.34 (-20.81; 6.10)		6.80 (-1.65; 15.97)
Fully adjusted model*	-9.18 (-27.93; 14.46)		5.24 (-5.06; 16.66)
Best model†	-2.06 (-19.02; 18.46)		5.48 (-3.67; 15.51)
Warm season			
Unadjusted model	-6.53 (-30.45; 25.62)	-6.44 (-19.48; 8.71)	9.61 (-1.36; 21.80)§
Fully adjusted model*	8.46 (-27.95; 63.28)	-2.42 (-15.19; 12.28)	6.16 (-6.77; 20.88)
Best model†	15.68 (-20.02; 67.33)	-2.18 (-14.49; 11.90)	6.83 (-4.43; 19.42)

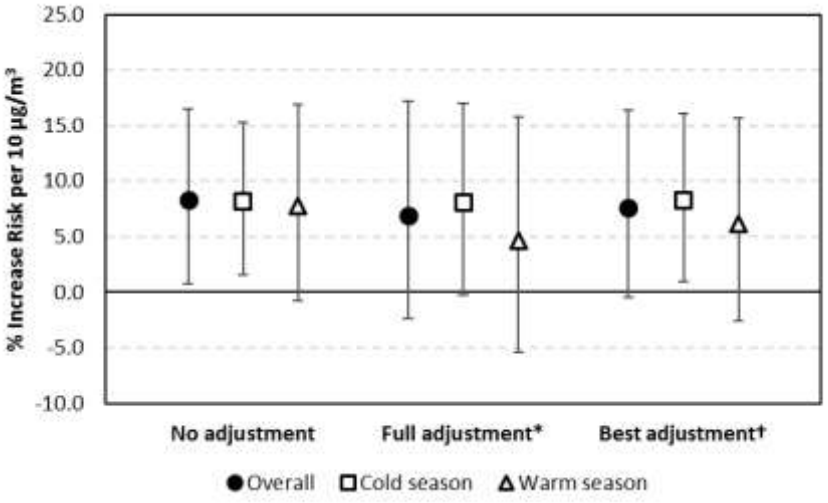
* Adjusted for area-specific proportion of males, mean age, average income, temperature and main use of the land

† For PM_{10} and NO_2 adjusted for area-specific mean age and main use of the land, for O_3 adjusted for area-specific proportion of males

‡ p-value <0.05

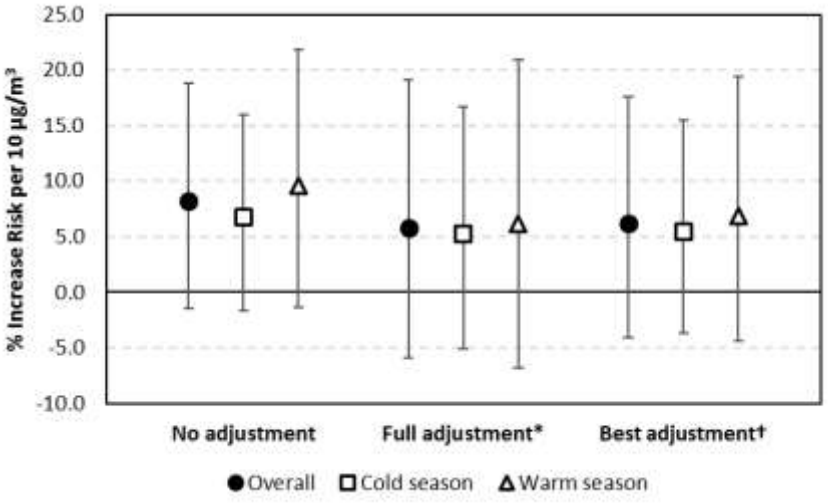
§ p-value <0.1

Figure S9. Generic Case Definition. Estimated % change in IPF incidence rate, with 95% Confidence Interval, for a 10 μg/m^3 increase in the average NO_2 concentration.



* Adjusted for area-specific proportion of males, mean age, average income, temperature and main use of the land
† Adjusted for area-specific mean age and main use of the land

Figure S10. Narrow Case Definition. Estimated % change in IPF incidence rate, with 95% Confidence Interval, for a 10 μg/m^3 increase in the average NO_2 concentration.



* Adjusted for area-specific proportion of males, mean age, average income, temperature and main use of the land
† Adjusted for area-specific mean age and main use of the land

Figure S11. Age and gender adjusted incidence rates. Comparison between areas with an average NO_2 concentration above and below 40 µg/m^3.

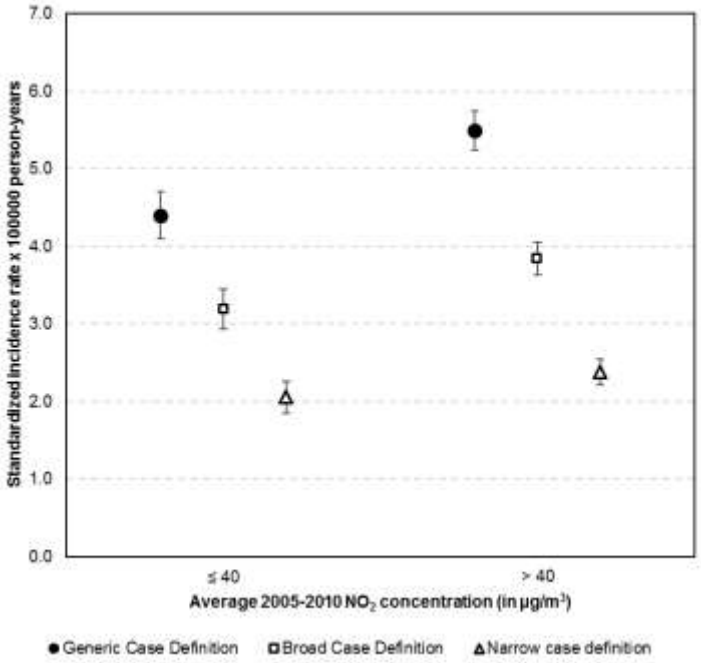


Figure S12. Age and gender adjusted incidence rates. Comparison between areas with an average NO_2 concentration above and below 20 µg/m^3.

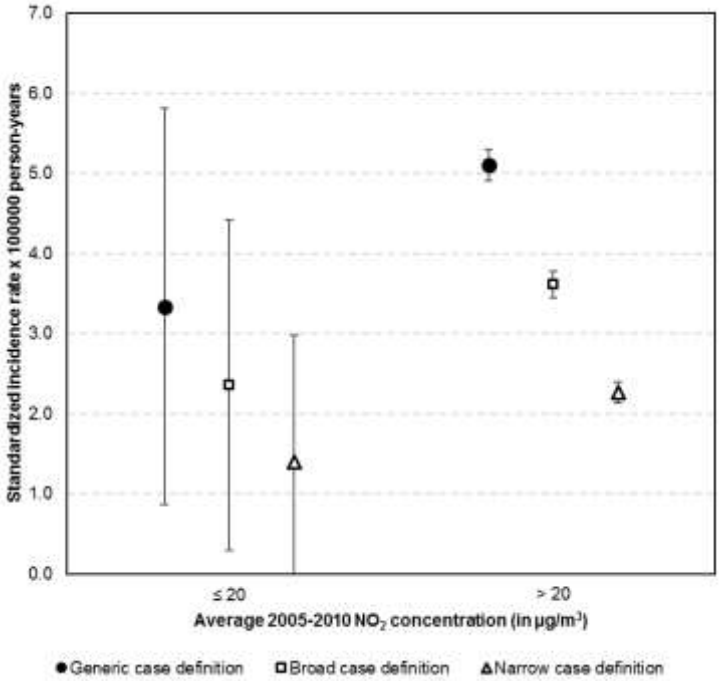
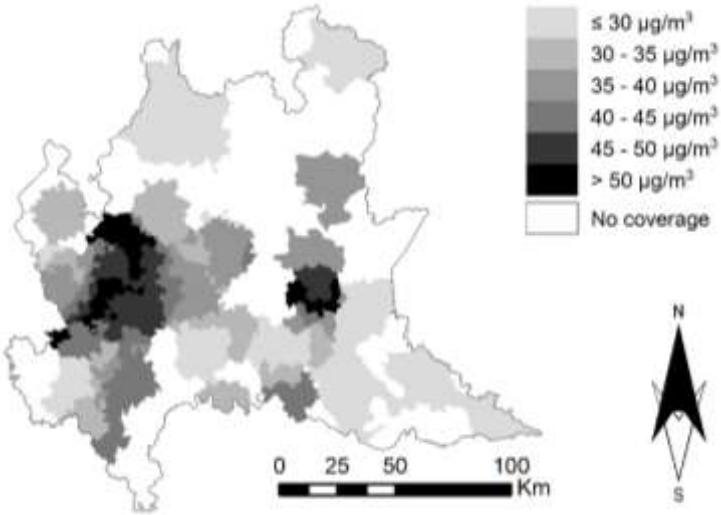


Figure S13. Map of the estimated daily average NO_2 concentration (in µg/m^3), based on background monitors, for the period 2005-2010 at a municipality level.



Shapefile of the Italian administrative regions and municipalities downloaded from the website of the Italian national institute of statistics (http://www.istat.it/it/archivio/24613).