Greenspace exposure and cancer incidence: A 27-year follow-up of the French GAZEL cohort

Mohammad Javad Zare Sakhvidi, Jun Yang, Jack Siemiatycki, Payam Dadvand, Kees de Hoogh, Danielle Vienneau, Marcel Goldberg, Marie Zins, Emeline Lequy, Bénédicte Jacquemin

Greenspace exposure and cancer incidence: a 27-year follow-up of the French GAZEL cohort

Mohammad Javad Zare Sakhvidi[1], Jun Yang[1], Jack Siemiatycki[2], Payam Dadvand[3,4,5], Kees de Hoogh[6,7], Danielle Vienneau[6,7], Marcel Goldberg[8,9], Marie Zins[8,9], Emeline Lequy[8], Bénédicte Jacquemin[1]

1. Univ. Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-35000 Rennes, France

2. Centre de recherches du centre hospitalier de l'université de Montréal, Québec, Canada

3. ISGlobal, Doctor Aiguader 88, 08003 Barcelona, Spain

4. Universitat Pompeu Fabra (UPF), Doctor Aiguader 88, 08003 Barcelona, Spain

5. CIBER Epidemiología y Salud Pública (CIBERESP), Melchor Fernández Almagro, 3-5, 28029 Madrid, Spain

6. Swiss Tropical and Public Health Institute, Basel, Switzerland

7. University of Basel, Basel, Switzerland

8. Inserm UMS 11, Villejuif, France

9. Université de Paris, Paris, France

Corresponding authors' email and adress: benedicte.jacquemin@inserm.fr Inserm UMR 1085; 9 avenue du Prof. Léon Bernard; 35000 RENNES

Competing Financial Interests: The authors declare they have no actual or potential competing financial interests.

Abstract

Background: Greenspace exposure has been suggested to be associated with a range of health outcomes. The available evidence on the association of this exposure with cancer is still very scarce and inconsistent.

Objectives: We aimed to study the association between greenspace exposure and all-site and site-specific (prostate, breast, colorectal, bladder, lung, and malignant melanoma of skin) cancer incidence in the GAZEL cohort.

Methods: This study was based on over 27 years of follow-up (1989-2016) of 19,408 participants across France. We assessed the residential greenspace exposure within several buffers as well as residential proximity to green spaces (agricultural, urban, and forests) in each follow-up. We used time-dependent Cox models, controlling for time-varying personal and area-level variables, with different lags between exposure and outcome. Additional analysis was conducted according to the urban-rural residence of the participants' over follow-up.

Results: Over the 294,645 person-years of follow-up, we registered 4,075 incident cases of cancer. We found an increase in the risk for all-sites cancer with an inter-quartile range increase of Normalized Difference in Vegetation Index across different buffers (hazard ratio (HR) of 1.08; 95% CI: 1.02, 1.14 for the 100m buffer). We found a positive association of all-sites cancer with proximity to agricultural lands (HR: 1.03; 95% CI: 1.00, 1.05), and forests (HR:1.04; 95% CI: 1.00, 1.07), but not with urban green spaces. The cancer site-specific analyses suggested a protective role of greenspace for breast, lung, and colorectal cancers (e.g. breast cancer HR at 100m buffer: 0.82; 95% CI: 0.69, 0.99). Non-significant associations were observed for prostate, bladder, and skin cancer. Stratified analyses based on urban, semi-urban, and rural classification did not suggest any differential pattern.

Conclusion: We identified an increased risk of all-site cancer with increased greenspace and proximity to agricultural lands and forests; whereas potential protective role of greenspace for breast cancer.

Keywords: Cancer incidence; Greenspace; Environmental health; Longitudinal study

1. Introduction

More than half of the world's population is already living in urban areas (in 2014, 48% in the less developed regions, and 78% in the more developed countries), and this proportion is continuously increasing (United Nations Population Division 2014). Urbanization leads to diminished human contact with the natural environment and increased exposure to anthropogenic pollutants. Natural environments such as greenspace (i.e. land at least partly covered with vegetation) can mitigate part of the adverse health effects of urbanization imposed on the human (Markevych et al. 2017; Twohig-Bennett and Jones 2018). Emerging evidence on the association between exposure to the natural environment, and more specifically greenspace, suggest potential beneficial effects of greenspace on human mental and physical health and well-being (Groenewegen et al. 2006; Markevych et al. 2017). The proposed pathways underlying such benefits are suggested to be improved social cohesion and physical activity, reduced stress, and decreased exposure to environmental pollutants such as air pollution, noise, and heat (Markevych et al. 2017).

A limited body of evidence has evaluated a potential association of greenspace exposure with the risk of cancer; however, the results have remained inconsistent. While some studies have reported a protective association of this exposure with cancer (Datzmann et al. 2018; O'Callaghan-Gordo et al. 2018),

others have reported an increased risk of cancer associated with this exposure (Astell-Burt et al. 2014; Carles et al. 2017; Richardson et al. 2010). More specifically, higher exposure to greenspace has been reported to be associated with reduced risk of non-melanoma skin cancer (Datzmann et al. 2018), breast cancer (O'Callaghan-Gordo et al. 2018), mouth and throat cancer (Datzmann et al. 2018), and prostate cancer (Demoury et al. 2017). In contrast, other studies have reported an increased risk of skin and brain cancer (Astell-Burt et al. 2014; Camille et al. 2017) or null associations (Datzmann et al. 2018). In this context, proximity to agricultural lands has been reported to be associated with an increased incidence of brain cancer (Camille et al. 2017). But these heterogeneous findings were reported by scarce studies with generally small populations and cross-sectional or ecological designs that had limited capability in studying possible links.

In this study, we assessed whether exposure to greenspace (defined as residential surrounding greenspace and proximity to specific types of greenspace) was associated with all-site and site-specific cancer incidence in the GAZEL cohort, a large and well-established national cohort with 27 years of follow-up data including a wide variety of personal and area-level covariates.

2. Materials and methods

2.1. Study design and population

GAZEL cohort is a prospective study of the French national electricity and gas company (Électricité de France-Gaz de France) workers, comprised of 20,625 individuals (15,011 men, 5,614 women) at baseline (aged at enrolment in the range of 35 to 50; mean= 44.21 years) (Goldberg et al. 2006; Goldberg et al. 2014). Participants were enrolled in 1989 and were followed-up annually by a) the mailed questionnaire at each year, b) regular data extraction from the files of the personnel records of employee up to retirement, c) data from medical departments of company up to retirement, and d) from national medical registries (Goldberg et al. 2014). In this analysis, we included the follow-ups to 2016. In addition to the demographic, socio-occupational, lifestyle, and occupational exposure data for each participant, data on health outcomes including cancer incidence were collected in each follow-up by questionnaires, medical examinations, and linkage to national health databases. Between 1989-2006 the postal codes of the participants were collected at each follow-up. From 2006 the complete residential address at each follow-up has been collected. Moreover, a residential history questionnaire was sent to the participants in 2006 to collect residential addresses before 2006. The responses were geocoded by the French postal company. For those who died before 2006 and had only a postal code, the residential addresses were located in the middle of the commune. The study area and geographical spread of the GAZEL cohort participants at the time of enrolment are shown in **Figure 1**. Contextual

indicators for each participant such as area-level deprivation index (for the year 2009) and time-varying population density of residential areas were linked to the residential addresses. We excluded participants who lived for over 20% of the follow-up time outside mainland France. Finally, we included 19,408 participants in our study (**Figure 2**). The GAZEL cohort protocol was approved by the French authority for data confidentiality and by the Ethics Evaluation Committee of the "Institut National de la Santé et de la Recherche Médicale" (INSERM) (IRB0000388, FWA00005831).

2.2. Greenspace exposure

We used residential proximity to the different types of greenspace (i.e. agricultural lands, urban greenspace, and forests), and residential surrounding greenness (applying a satellite-based index) as the measures of exposure to greenspace.

2.2.1. Residential proximity to green spaces

The land cover data were downloaded from the European CORINE land cover inventory, available for the years 1990, 2000, 2006, and 2012 (Copernicus Land Monitoring Service 2020). We used CORINE definition for "artificial, non-agricultural vegetated areas" as "*urban green spaces*"; "arable land, permanent

crops, and pastures" as "*agricultural lands*"; and "forests, scrub and/or herbaceous vegetation associations" as "*forests*" (Copernicus Land Monitoring Service 2020). We calculated the Euclidean distance from each residential address (as a point) to nearest distances to different land covers (as a polygon) of the closest date of available land cover data by NNjoin plugin for QGIS (Tveite 2017). The inverse of the distances (as 1/km) was defined as "*residential proximity*" to each type of greenspace.

2.2.2. Residential surrounding greenspace

We extracted residential surrounding greenspace from the normalized difference vegetation index (NDVI) derived from the remote sensing data of Landsat satellites (The United States Geological Survey 2020) for each year between 1989 (as the earliest year for which it was possible to obtain participants epidemiologic data) up to the time of cancer incidence or censoring (up to 31st of December 2016). NDVI is a unitless greenspace indicator (ranges between −1 and 1) based on land surface reflectance of visible (red) and near-infrared parts of the electromagnetic radiation (Bannari et al. 1995). It is computable based on surface reflectance data, with an image available every16 days and at a spatial resolution of 30m x30m via Landsat satellites. The data were acquired from Landsat 5 for 1984-2012, Landsat 7 for 2013, and Landsat 8 for afterward. NDVI was calculated based on red and infra-red bands [NDVI= (Near infra-

red)/ (Red- Near infra-red)] (bands of 3, 4 for Landsat 5 and bands 4 and 5 for Landsat 7 and 8)(The United States Geological Survey 2017). Images acquisition and NDVI calculations were performed by using the Google earth engine platform (Gorelick et al. 2017). We used atmospherically corrected and cloud, shadow, water, snow, and per-pixel saturation masked rasters (Google Earth Engine 2020).

In this study we used the NDVI values at the greenest time of each year as an indicator of greenspace exposure of that year to maximize the contrast in our exposure; this period also had the least problems with cloud coverage interference (Vienneau et al. 2017). To select the month(s) with the highest greenspace exposure, the greenspace of mainland France for different months of the year was calculated (Temam et al. 2017). The highest greenspace exposure was observed between May-July. Therefore, we selected the May-July timeslot, then selected the image with the lowest cloud coverage at this timeslot, and calculated the NDVI values.

Gong et al. found that the direct environment around residential addresses could be described at best with a 500m buffer especially for older adults (Gong et al. 2014). Additionally, smaller buffer sizes have been reported to have a stronger association with perceived health (possibly due to the visual impact of greenspace) (Su et al. 2019). Therefore, to consider possible different causal

mechanisms of greenspace on the incidence of cancer, we used four different buffer sizes (at 100, 300, 500, and 1000m radius), and extracted corresponding average NDVIs for each residential address annually.

All NDVI values below zero (due to the presence of water bodies or too high cloud coverage) were converted to a no-data pixel and then were imputed with the average of NDVI at the same location of the other available years.

2.3. Outcomes

The outcomes in this study were all-sites and most prevalent site-specific cancers including prostate cancer (ICD-10 code: C61), breast cancer (ICD-10 code: C50), colorectal cancer (ICD-10 codes: C18 for colon cancer; C19 for rectosigmoid cancer; C20 for rectum cancer), bladder cancer (ICD-10 code: C67), lung cancer (ICD-10 code: C34) and malignant melanoma of skin (ICD-10 code: C43). During the employment period of participants, cancer records were collected by the Électricité de France- Gaz de France (EDF-GDF) medical department (except non-melanoma skin cancers). After retirement, the cancer validation platform (PRIMEV'R), and since 2007, SNIIRAM database (Système National d'Information Inter-Régimes de l'Assurance Maladie) were used for cancer incidence data collection (French health insurance 2019). PRIMEV'R platform was developed in 2008 to collect cancer incidence data after the

retirement period of participants. PRIMEV'R is based on response to the annual questionnaire by cohort participants. The SNIIRAM is the national health administrative database that records all data on the use of the health system leading to reimbursement; it includes two main sources of information for cancer: hospitalization (diagnoses ICD-10 coded), and chronic diseases register (including cancers coded with ICD-10). In all cases the criteria were the date of cancer diagnosis or report. We excluded participants with cancer at the baseline (n= 5).

2.4. Covariates

Individual level covariates were collected annually from the inclusion and follow-up questionnaires and included sex, age at enrolment, body mass index (BMI in kg/m2; weight was available annually except for 1989 and height was available annually except for 1989, 1991-1996), family status (in a couple or not), education (categorized in five groups including 9-11 years of education, 12-13 years of education, 14-15 years of education, other secondary diploma, and other diploma), smoking status (smoker or non-smoker), smoking intensity (cumulative pack-years as a continuous variable), passive smoking at home or at work (yes or no; reported two times in 1990 and 1996), alcohol consumption (abstinent; light drinker; moderate drinker; heavy drinker; and unclear pattern), fruit or vegetable use (never or less than once a week; once or twice a week; more than twice a week but not every day; every day or almost every day;

reported four times in 1998, 2004, 2009 and 2014), exposure to selected carcinogens based on occupational exposure to a list of selected proven carcinogens according to the International Agency for Research on Cancer (IARC) classification (cadmium, polychlorinated biphenyls, asbestos, trichloroethylene, benzene, aromatic solvents, pitch, silica, hexavalent chromium, carbon gasification) determined by a job exposure matrix (Imbernon et al. 1991) (categorized in four groups, including 0: no exposure to occupational carcinogens; 1: exposure to one occupational carcinogen; 2: exposure to two occupational carcinogens; 3: exposure to three and more occupational carcinogens), socio-occupational status from baseline (executive; supervisor; worker; available annually). We also assigned exposure to airborne fine particulate matter with aerodynamic diameter <2.5 μm ($PM_{2.5}$) at the individual level. We used the data from a previously developed land-use regression (LUR) model for the whole of western Europe including France (De Hoogh et al. 2018). The LUR was based on the 2010 AirBase database (maintained by the European Environmental Agency), chemical transportation models, and satellite data. The model was back and forward extrapolated from 1989 to 2015 using annual correction factors based on the Danish Eulerian Hemispheric Model (DEHM) at level 1 of NUTS (nomenclature of territorial units for statistics; NUTS is a hierarchical system for dividing up the economic territory of the European Union and the United Kingdom (European Union 2020)). Finally, cumulative exposure of the participants to $PM_{2.5}$ from 1989 until

the time of cancer incidence or censoring was considered as the measure of participants' exposure to air pollution. Distance to the nearest major roads (as a logarithmic scale) for each residential address was also included in the models as a proxy of traffic-related pollutants.

Data on area-level variables including neighborhood deprivation based on French deprivation index categorized by tertile (was available only for 2009) (Rey et al. 2009), the time-varying population density of the residential area (for urban, semi-urban and rural classification based on French National Institute of Statistics and Economic Studies) were also collected (Rey et al. 2009). The French deprivation index is based on four variables including median household income, percentage of high school graduates, percentage of blue-collar workers, and unemployment rate (Temam et al. 2017).

2.5. Statistical analysis

2.5.1. Main analyses

We developed time-dependent Cox proportional-hazard regression models, with all-site as well site-specific cancer incidence as the outcome variable (one at a time in separate models), each indicator of greenspace exposure as the exposure variable (one at a time in separate models), and age at baseline as the time scale (Fisher and Lin 2000). We did a site-specific analysis for selected cancer sites.

For site-specific analyses, we excluded cases with other cancers (for example for the analysis of breast cancer as an outcome, we kept breast cancer cases and all those without any cancer). Analyses on prostate and breast cancer were conducted only among men and women, respectively.

The latency between initial tumor growth following exposure to environmental hazards and the clinically detectable state of different categories of cancer remains unclear. Therefore, in the main analyses we considered a 10-year delay between exposure and outcome as proposed by previous studies (Rothman et al. 2008), following Equation 1:

$$h\ \left(t, X(t)\right) = h_0(t) exp\left[\sum_{i=1}^{p_1} b_i X_i\ + \sum_{j=1}^{p_2} \delta_j X_j (t - L_j)\right] \text{ Equation (1)}$$

where h_g is the hazard ratio for exposure to "greenspace", b_i is the regression coefficient for the time-independent variables; δ_j is the regression coefficient for the time-dependent variables; Lj denotes time delay (in this case we used 10 years cumulative exposure for PM2.5); Xi and Xj are time-independent and time-dependent variables, respectively; t is the time and h0 is the baseline hazard function. We entered 10-years moving average of greenspace exposure measures, exposure to outdoor air pollution (cumulative exposure to PM2.5), distance to major roads, BMI, smoking status, smoking intensity, passive smoking, familial status, frequency of using fruits or vegetables, alcohol

consumption, and urban classification as the time-dependent variables in the main models (and five years, and without lag situation for sensitivity analyses). All main models also contained personal socio-occupational status, occupational exposure to carcinogens, education, age at baseline, and French deprivation index as the time-independent variables.

The associations were presented in the form of hazard ratio (HR) and corresponding 95% confidence intervals (95% CIs). To consider the sex-difference in the incidence rate at baseline, we used the sex-stratified extended Cox model.

The percentage of missing values of the selected variables was between 0 and 37.5% (for smoking intensity). We used longitudinal multiple imputations (using mice (Buuren and Groothuis-Oudshoorn 2010) and miceadds (Robitzsch et al. 2016) packages in R) to generate and analyze five imputed datasets. Participants identifier was considered as a level-2 imputation parameter. The density plots of imputed variables were visualized for checking the acceptance of imputation. The models were run on multiple imputed datasets and corresponding parameters were estimated in each imputed dataset separately. Estimates were combined using Rubin's rules and pool function.

2.5.2. Stratified analyses

We performed stratified analyses on all-sites and selected site-specific cancers for urban, semi-urban, and rural classification of the residential area of participants. Allocation of participants to urban, semi-urban, or rural residents was based on the longest duration of residence in the specific classification throughout the follow-up (in terms of population density). Rural residents generally were living closer to agricultural lands than urban green space (for urban residents we observed the contrary). Therefore, we did not consider proximity to urban green space for rural residents and proximity to agricultural lands and forests for urban residents.

2.5.3. Sensitivity analyses

Several sensitivity analyses were considered for all-sites, prostate, breast, and colorectal cancers (we did not do sensitivity analysis for other site-specific cancer due to too few cases in each stratum). At first, we did sensitivity analysis based on occupational exposure to carcinogens. Furthermore, tested the associations for different lags between the exposure and outcomes. We also did a sensitivity analysis by excluding participants with residential mobility. We defined a mobility variable according to change in residential location of participants during the follow-up, categorizing participants into movers (those with at least one change in the geocoded residential address) and non-movers (those who lived in the same location during all the follow-up).

All HRs in this study are reported based on one interquartile range (IQR) increase of exposure. We *a priori* selected NDVI 100m as the main surrounding greenspace metrics but also reported the results for other buffer sizes (200, 500, 1000m) in the supplement. The HRs of residential proximity to agricultural lands, forests, and urban greenspace is reported per one km proximity. Results are reported first for all-sites then site-specific cancers ordered based on the number of cases. For Cox proportional hazard regression, we used the R package "survival" (Therneau 2000, 2019). All analysis was conducted in R version 3.5 (R Development Core Team).

3. Results

The study population included 19,408 participants (mean age at the enrolment: 43.74±3.50 years); mainly male (72.90%) and non-smoker (81.50%) (**Table 1**). By the end of 2016 (the last follow-up for this analysis), the remaining participants were aged 63–77 (mean=71.05), and almost all of them were retired. **Tables S1** describe the general characteristics of participants according to their mobility.

Over the 294,645 person-years, we observed a total of 4,075 cases of cancer. Of those 1401 (34.38%) were prostate, 378 (9.28%) breast, 367(9.01%) colorectal,

268 (6.58%) bladder, 237 (5.82%) lung, and 156 (3.83%) malignant melanomas of the skin. The mean baseline exposure to surrounding greenspace (based on NDVI within 100m) was 0.47±0.17 and 0.46±0.17 for cancer and non-cancer groups respectively (p=0.056). The NDVI values in the different buffer sizes were highly correlated (Pearson's correlation coefficients > 0.80) (**Table 2**). Participants with cancer were significantly closer to the forests, but farther from urban green spaces compared to the non-cancer group at the baseline (**data are not shown**). A negative correlation was found between $PM_{2.5}$ and NDVI in all buffer sizes (Pearson's correlation: -0.55; 95% CI: -0.56: -0.54 for NDVI at 100m) (**Table 2**). Proximity to agricultural lands and forests were also correlated with lower $PM_{2.5}$; however, proximity to urban green space was correlated with increased $PM_{2.5}$ (**Table 2**). **Table S2** shows exposure to greenspace in urban, semi-urban, and rural areas.

Table 3 shows the associations between exposure to greenspace (per an IQR increase in exposure) and cancer incidence (results of 300, 500, and 1000m buffer sizes are presented in **Table S3**). We found a hazard ratio of 1.075 for all-site cancers per an IQR increase in NDVI at 100m buffer size (95% CI: 1.016, 1.137). A significant increase in the risk of all-site cancers was observed for the proximity to agricultural lands (HR: 1.026; 95% CI: 1.002, s1.052) and forests (HR:1.038; 95% CI: 1.004, 1.074). The result for NDVI at other buffer sizes and the proximity to the urban green spaces were inconclusive.

In site-specific analyses, we found mainly protective associations between residential surrounding greenspace and breast cancers; however, the associations attained statistical significance only for the small number of associations. The strongest protective association was observed for breast cancer in relation with the residential surrounding greenness at 100m buffer size (HR: 0.824; 95% CI: 0.687, 0.989), followed by colorectal cancer at 1000m buffer size (HR: 0.860; 95% CI: 0.690, 1.071). In contrast, higher surrounding greenness was associated with a higher significant risk of prostate cancer (HR: 1.148; 95% CI: 1.042, 1.265 for NDVI 100m). Proximity to different types of green space (agricultural lands, urban green spaces, and forests) was not associated with a significant increase or decrease in the risk of selected cancer sites.

We performed a stratified analysis based on the urban classification (**Table 4** and **S2**). Higher residential surrounding greenspace was suggestive for the reduction in the risk of breast and colorectal cancer incidence in all three areas. Proximity to agricultural lands in rural areas was also suggestive of a non-significant decrease in the risk of breast cancer incidence (HR: 0.885; 95% CI: 0.781, 1.002). As for the whole population, the increased risk of all-site cancer was observed in all three areas (urban, semi-urban, and rural); however, the results were mainly significant in semi-urban areas.

The sensitivity analyses limited to participants without exposure to occupational carcinogens (only for all-site, prostate, breast, and colorectal cancers) showed a similar pattern (direction and magnitude of the estimates) as the main analyses (**Table S4**). However, most of the observed associations for prostate cancers became non-significant (e.g. HR: 1.153, 95% CI: 0.99:1.34 for NDVI 100m) probably because occupational exposure could be a driver for cancer outcomes, and also the smaller number of observations and lower statistical power. Changing the lag time in the models showed a sensitivity of breast cancer results to the selection of lag period. With changing lag period from 10 years to 5 years, and also without lag, the observed protective associations for breast cancer became weaker and lost their statistical significance (**Tables S5** & **S6**). But for all-sites, skin and lung cancer, the findings of this sensitivity analysis showed a similar pattern (in terms of the direction of observed association and observed significant associations) as the main analyses. Excluding participants who moved, we found a significantly higher risk for all-sites cancer for proximity to forest, urban green space, and surrounding greenness at 100 and 300m buffer sizes (**Table S7**). Proximity to agricultural lands was significantly associated with a lower risk of colorectal cancer after excluding participants who had moved during the study period.

4. Discussion

In this study, we examined the association between exposure to greenspace (characterized as the residential surrounding greenness and proximity to specific types of green spaces) and all-site and site-specific cancer incidence during 27 years of the follow-up of the GAZEL cohort participants. While higher residential surrounding greenness was suggestive of reduced risk of breast, and lung cancer, it was also associated with an increased risk of all-sites as well as prostate cancer. We also observed an increased risk of all-site cancers with the increasing residential proximity to forests and agricultural lands, but not urban green spaces. Our findings remained consistent after we stratified our analyses based on the degree of urbanity, or sensitivity analysis by excluding those who moved or were exposed to selected occupational carcinogens. This study by considering a wide range of covariates (which are measured year by year), different types of green space exposure, and different outcomes on a large cohort of participants with near 30 years of follow-up adds to the available evidence on the link between green space and cancer. The main message of our study is: part of this heterogeneity would be due to the exposure assessment method, not the inclusion of different confounders, combining different exposure types, and the power of the studies.

A limited number of studies reported possible associations between exposure to greenspace and cancer incidence, and are inconsistent depending on the site of cancer, type of greenspace, the measure of exposure, and also the geographical

location of the study (Astell-Burt et al. 2014; Datzmann et al. 2018; Demoury et al. 2017; Iyer et al. 2020; O'Callaghan-Gordo et al. 2018). Negative associations were reported between mortality from esophageal cancer and exposure to green space when using winter greenspace instead of summer greenspace (Wu et al. 2008). Consistent with our findings, a study in Australia (Astell-Burt et al. 2014) found an increase in the risk of skin cancer prevalence with an increased percentage of public green space around the residence area. Another study conducted in France reported an increased risk for proximity to agricultural lands and brain cancer (Carles et al. 2017).

The observed association within and between the studies is heterogeneous. We found only four cohort studies on the association between greenspace and cancer (Datzmann et al. 2018; Iyer et al. 2020; James et al. 2016; Klompmaker et al. 2020). Only two of these studies were on cancer incidence (Datzmann et al. 2018; Iyer et al. 2020). And from these two studies, one of them was a semi-individual cohort (Datzmann et al. 2018), and another one was only on lethal prostate cancer cases (n= 898) (Iyer et al. 2020). Therefore, to the best of our knowledge, our study is among the few available longitudinal studies on the association between greenspace and cancer incidence, which would add to the previous understanding about the association between greenspace and cancer. Despite the higher quality of the available cohort studies on the association between green space and cancer (Datzmann et al. 2018; Iyer et al. 2020; James

et al. 2016; Klompmaker et al. 2020) compared to the available cross-sectional or time-series studies (Richardson et al. 2013 & 2010; Bixby et al. 2015; Kim et al. 2019; Mitchell et al. 2008; Wu et al. 2008), all of them reported heterogeneous results across different subgroups. For example, in a cohort study in the Netherlands (Klompmaker et al. 2020), the surrounding greenspace was not associated with lung cancer mortality, however, the findings for non-elderly subjects were different from the elderly subjects. Lyer et al. (2020) in the USA found a 10% lower risk of lethal prostate cancer associated with greenspace exposure for high-density areas (95% CI: 0.82-0.99), whereas they found an 11% increase in risk for participants in low-density areas (95% CI: 0.95-1.29). Datzman et al. (2018) in Germany found that greenspace was significantly associated with a 4% decrease in breast cancer, a 16% decrease in non-melanoma skin cancer, and an 11% decrease in mouth and throat cancers in all subjects. However, subgroup analysis based on sex in their study showed an increased risk of colorectal, mouth and throat, and non-melanoma skin cancer in men. In addition to the heterogeneous results across different groups, the use of different exposure measures in the same study also has given controversial associations from significantly protective to significantly risk factor associations (O'Callaghan-Gordo et al. 2018). In a Spanish case-control study, the presence of urban greenspace was associated with a lower risk of breast cancer, but the presence of agricultural greenspace in proximate buffer (300 m), and not distant

buffer (300-500 m) was associated with a significant increase in the risk of breast cancer (O'Callaghan-Gordo et al. 2018).

In addition to the within studies heterogeneity of the findings, the findings for the same outcomes across the studies are also controversial. In one case-control study, O'Callaghan-Gordo et al. (2018) found an increase in the risk of breast cancer with an increase in NDVI at 300m buffer, whereas Datzman et al.(2018) found a protective association with NDVI as a measure of greenness(RR:0.96; 95% CI: 0.92: 0.99).

Cancer risk could be impacted by many factors, and our findings on the association between greenspace exposure and cancer should be interpreted with caution. However, we have several explanations for the observed increased cancer risk in areas with higher greenspace. In general, areas with higher greenspace have lower road density and consequently lower traffic, therefore lower levels of traffic-related air pollutants (Su et al. 2011), but higher concentrations of ozone, biogenic volatile organic compounds, and secondary organic aerosols (Su et al. 2011). Trees can decrease the spread of traffic air pollutants and therefore increase the air pollution concentration in the roads (Markevych et al. 2017). We found a positive correlation between proximity to urban greenspace and $PM_{2.5}$, which likely reflects the generally higher air pollution exposure in urban areas. A study in the USA also reported higher

levels of NO_2 and $PM_{2.5}$ in park-adjacent neighborhoods compared to the parks themselves or the entire region (Su et al. 2011). Higher surrounding greenspace is thought to be associated with increased outdoor physical activity (Richardson et al. 2013). On the other side, a longer time of outdoor activities is associated with higher exposure to ultraviolet radiation and possibly increased exposure to traffic-related air pollutants (Grigsby-Toussaint et al. 2011). For all-site cancers, our findings suggest an increase in the cancer risk with proximity to agricultural lands and forest. Proximity to the agricultural lands has been thought to be associated with an increase in exposure to pesticides (Deziel et al. 2015), among which some could increase the risk of cancer (IARC 2015). In agreement with this hypothesis, O'Callaghan-Gordo *et al.* (O'Callaghan-Gordo et al. 2018) also found a positive association between the residential proximity to agricultural lands and breast cancer. Our observed association for the proximity to forest could have been, in part, due to its high correlation with distance to agricultural lands. As we found in correlation analysis, those areas which were close to the forests, were also closer to the agricultural lands. In the case of proximity to green space areas, we used CORINE land cover data. Better resolution land use data at the national level, such as Urban Atlas for cities are not yet available. Despite its wide application in different publications, which increases the comparability of our results with other studies, it is relatively coarse and would not detect greenspace, such as private gardens or small public green spaces smaller than 25 hectares (Bossard et al. 2000; Su et al. 2011). Our null finding

for the association between the residential proximity to urban green spaces and all-site and site-specific cancer could be attributable to this problem (Annerstedt et al. 2012).

Our finding on the different associations between movers and non-movers also replicated in another study on the association between greenspace and prostate cancer incidence (Iyer et al. 2020). It seems that mechanisms related to the association between green space and caner are related to environmental context. Additionally, we found that the non-movers were most likely to live in the semi-urban areas, had higher exposure to occupational carcinogens, but lower education, higher distance to urban greenspace, lower distance to the nearest forest and agricultural lands. This could be interpreted that the non-movers in this study were those with higher exposure to known risk factors of cancer, which can partly explain our findings on the stronger association for urban green space with non-movers in this study.

We found a stronger protective association for breast cancer at 10 years lag results compared to the 5 years and without lag. If confirmed, this finding would be of high importance especially for understanding the critical risk period for the exposure-outcome association. It means that the results are sensitive to the latency period; further exploration for potential risk period is necessary.

Because of a small number of participants in site-specific analyses, most of the associations did not reach statistical significance. Therefore, our findings for breast, colorectal, and lung cancer should be interpreted with caution.

Strengths and limitations: Our study is one of the few prospective cohort studies looking for the possible association between greenspace exposure and cancer incidence. Furthermore, we have a long follow-up time and yearly updates of the exposure data. Our results are based on more than 19,000 participants and 4,000 incident cases of cancer over a 27-year follow-up period. Participants in this study were from different socio-occupational groups residing in all regions of mainland France with different deprivation indices. These characteristics could increase the generalizability of our findings, even if our participants tend to be healthier (possible healthy workers effect), and wealthier than the general population. Additionally, we assessed residential surrounding greenness for each year of follow-up using annual satellite greenspace data. This extensive characterization of exposure was an advancement compared to most of the previous studies that mainly relied on greenspace data at one point in time or applied aggregated data for estimating greenspace exposure. For the distances, we used the inverse of the distance to the green spaces in the models. We choose this function based on hypotheses on linear increase/decrease of the effect of greenspace with distance, which is used commonly in related papers. However, there are several other shapes, such as the inverse of the square distance which

would be interesting to test in future studies. Considering the institutional follow-up of the participants during their working years and the robust data linkage after their retirement, there was very a small loss to follow-up for outcomes in the GAZEL cohort over about three decades of follow-up. Availability of yearly follow-up data on both exposures and covariates made it possible for us to enter the covariates as a time-varying variable in the models. We also included $PM_{2.5}$ and distance to the road to control for confounding by air pollution and of traffic-related pollutants in our models. Our findings on different types of green space and also cancer-sites would be useful to describe the available heterogeneity across the studies. The results highlight that the potential association between greenspace and cancer could not be generalized and should be defined separately based on the type of the greenspace and also the cancer site. As we found a beneficial association mostly for breast cancer, and suggested negative association for prostate and all-site cancers. Additionally, our findings highlight that part of observed heterogeneity between studies would be due to the methods of exposure assessment, combining different green spaces, and the statistical power of the studies.

Our study faced several limitations. The major limitation in our study is limited statistical power for in-depth site-specific cancer analyses. All site-cancer results can be misleading as there are different mechanisms for site-specific cancers. However, it is important to note that most of the cancers included in the analyses

are solid tumors. Our results may be driven by prostate cancer results as they count for one-third of the cases. When excluding prostate cancer cases, all the estimates lost significance shifting to null estimate (data are not shown), supporting that hypothesis, even if we cannot rule out the power issue. Additionally, missing data forced us to do multiple imputations, with the inherent possible biases that could results(White and Carlin 2010), nevertheless, we choose the multiple imputation method which is one of the most recognized, with 10 datasets and resulting in an expected variable distribution(Carpenter and Kenward 2012).

Cancer risk could be impacted by many socioeconomic and environmental factors, and our findings on the association between greenspace exposure and cancer should be interpreted with caution, and the role of some other cofounders should not be neglected. For example, we had no data on all of the possible confounders and influential factors such as time-activity patterns, and cancer screening behavior of participants. However, we have tried to reduce the role of confounding in our study in different stages of the study including study design and analysis, and interpretation of the results. In the design stage of the study, we have tried to consider as much as possible different personal (behavioral and nutritional variables) and contextual variables in the models to consider possible confounding in the association. For example, we included detailed smoking history, passive smoking, occupational exposure to different carcinogens

extracted from a job exposure matrix, detailed fruit or vegetable use, physical activity, and history of cancer in the family members.

However, our cohort was not different in terms of personal and contextual variables at the start of the study. Additionally, participants in this study were among the workers of the same company, which would also additionally mitigate part of possible confounding because of socioeconomic status. We considered participants' home location as the point of exposure, overlooking the exposure that could happen in other microenvironments such as the workplace or its commuting route. Our characterization of greenspace exposure also did not include or assess the quality of these spaces. However, using NDVI in this study has increased the comparability of our results with other available studies which mostly used this index as an indicator of surrounding greenspace.

5. Conclusion

The results of this study suggest different associations between exposure to greenspace and cancer incidence based on cancer site and type of greenspace. We found an increased risk of all-site cancer with increased greenspace and proximity to agricultural lands and forests. We found suggestive results for a protective role of greenspace for breast, lung, and colorectal cancer. While many studies reported beneficial effects of greenspace on human health, our findings suggest that it is not generalizable to all outcomes. Our findings require

confirmation by other studies conducted in different climates and populations. Future studies should consider more precise greenspace exposure measures such as use, quality, and type of greenspace, while trying to shed light on the potential mechanisms underlying such an association.

Funding

This study was supported by the ARC Foundation which funded OCAPOL study (A longitudinal observatory of the effects of chronic exposure to outdoor air pollution on cancer risk) within the framework of the CANC'AIR call. Payam Dadvand is funded by a Ramón y Cajal fellowship (RYC-201210995) awarded by the Spanish Ministry of Economy and Competitiveness.

Role of the funding source

The study funder did not contribute to the study design, data collection, data analysis, data interpretation, or writing of this manuscript. The corresponding author had full access to all the data used in this study and had final responsibility for the decision to submit it for publication.

Contributors

MZ, JY, EL, and BJ contributed to the study design. MZ did the statistical analysis and prepared the draft. All authors contributed to the interpretation of data, provided critical revisions to the manuscript, and approved the final draft.

References

Annerstedt M, Östergren P-O, Björk J, Grahn P, Skärbäck E, Währborg PJBph. 2012. Green qualities in the neighbourhood and mental health–results from a longitudinal cohort study in southern sweden. 12:337.

Astell-Burt T, Feng X, Kolt GSJJECH. 2014. Neighbourhood green space and the odds of having skin cancer: Multilevel evidence of survey data from 267072 australians. 68:370-374.

Bannari A, Morin D, Bonn F, Huete AJRsr. 1995. A review of vegetation indices. 13:95-120.

Bixby, Honor, Susan Hodgson, Léa Fortunato, Anna Hansell, and Daniela Fecht. "Associations between green space and health in English cities: an ecological, cross-sectional study." PLoS One 10, no. 3 (2015): e0119495.

Bossard M, Feranec J, Otahel J. 2000. Corine land cover technical guide: Addendum 2000.

Buuren Sv, Groothuis-Oudshoorn KJJoss. 2010. Mice: Multivariate imputation by chained equations in r.1-68.

Camille C, Ghislaine B, Yolande E, Clément P, Lucile M, Camille P, et al. 2017. Residential proximity to agricultural land and risk of brain tumor in the general population. 159:321-330.

Carles C, Bouvier G, Esquirol Y, Piel C, Migault L, Pouchieu C, et al. 2017. Residential proximity to agricultural land and risk of brain tumor in the general population. Environmental research 159:321-330.

Carpenter J, Kenward M. 2012. Multiple imputation and its application:John Wiley & Sons.

Copernicus Land Monitoring Service. 2020. Corine land cover. Available: https://land.copernicus.eu/pan-european/corine-land-cover.

Datzmann T, Markevych I, Trautmann F, Heinrich J, Schmitt J, Tesch FJBph. 2018. Outdoor air pollution, green space, and cancer incidence in saxony: A semi-individual cohort study. 18:715.

De Hoogh K, Chen J, Gulliver J, Hoffmann B, Hertel O, Ketzel M, et al. 2018. Spatial pm2. 5, no2, o3 and bc models for western europe–evaluation of spatiotemporal stability. 120:81-92.

Demoury C, Thierry B, Richard H, Sigler B, Kestens Y, Parent M-EJEi. 2017. Residential greenness and risk of prostate cancer: A case-control study in montreal, canada. 98:129-136.

Deziel NC, Friesen MC, Hoppin JA, Hines CJ, Thomas K, Freeman LEBJEhp. 2015. A review of nonoccupational pathways for pesticide exposure in women living in agricultural areas. 123:515-524.

European Union. 2020. Statistical regions in the european union and partner countries. In: Statistical regions in the European Union and partner countries, (Union E, ed). Luxembourg.

Fisher LD, Lin DYJAroph. 2000. Time-dependent covariates in the cox proportional-hazards regression model. 20:145-157.

French health insurance. 2019. Sniiram. Available: https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/sniiram/finalites-du-sniiram.php.

Goldberg M, Leclerc A, Bonenfant S, Chastang JF, Schmaus A, Kaniewski N, et al. 2006. Cohort profile: The gazel cohort study. 36:32-39.

Goldberg M, Leclerc A, Zins MJIjoe. 2014. Cohort profile update: The gazel cohort study. 44:77-77g.

Gong Y, Gallacher J, Palmer S, Fone DJIJoBN, Activity P. 2014. Neighbourhood green space, physical function and participation in physical activities among elderly men: The caerphilly prospective study. 11:40.

Google Earth Engine. 2020. Landsat collections. Available: https://developers.google.com/earth-engine/datasets/catalog/landsat.

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D. 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. 202:18-27.

Grigsby-Toussaint DS, Chi S-H, Fiese BHJIjohg. 2011. Where they live, how they play: Neighborhood greenness and outdoor physical activity among preschoolers. 10:66.

Groenewegen PP, Van den Berg AE, De Vries S, Verheij RAJBph. 2006. Vitamin g: Effects of green space on health, well-being, and social safety. 6:149.

IARC. 2015. Iarc monographs volume 112: Evaluation of five organophosphate insecticides and herbicides. World Health Organization, Lyon.

Imbernon E, Goldberg M, Guénel P, Bitouze F, Brement F, Casal A, et al. 1991. Matex: Une matrice emplois-expositions destinée à la surveillance épidémiologique des travailleurs d'une grande entreprise (edf-gdf). 52:559-566.

Iyer HS, James P, Valeri L, Hart JE, Pernar CH, Mucci LA, et al. 2020. The association between neighborhood greenness and incidence of lethal prostate cancer: A prospective cohort study. Environmental Epidemiology 4:e091.

James, Peter, Jaime E. Hart, Rachel F. Banay, and Francine Laden. "Exposure to greenness and mortality in a nationwide prospective cohort study of women." Environmental health perspectives 124, no. 9 (2016): 1344-1352.

Kim, Sera, Honghyok Kim, and Jong-Tae Lee. "Interactions between ambient air particles and greenness on cause-specific mortality in seven Korean metropolitan cities, 2008–2016." International journal of environmental research and public health 16, no. 10 (2019): 1866.

Klompmaker, Jochem O., Gerard Hoek, Lizan D. Bloemsma, Marten Marra, Alet H. Wijga, Carolien van den Brink, Bert Brunekreef, Erik Lebret, Ulrike Gehring, and Nicole AH Janssen. "Surrounding green, air pollution, traffic noise exposure and non-accidental and cause-specific mortality." Environment international 134 (2020): 105341.

Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, et al. 2017. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. 158:301-317.

Mitchell, Richard, and Frank Popham. "Effect of exposure to natural environment on health inequalities: an observational population study." The lancet 372, no. 9650 (2008): 1655-1660.

O'Callaghan-Gordo C, Kogevinas M, Cirach M, Castaño-Vinyals G, Aragonés N, Delfrade J, et al. 2018. Residential proximity to green spaces and breast cancer risk: The multicase-control study in spain (mcc-spain). 221:1097-1106.

Rey G, Jougla E, Fouillet A, Hémon DJBph. 2009. Ecological association between a deprivation index and mortality in france over the period 1997–2001: Variations with spatial scale, degree of urbanicity, age, gender and cause of death. 9:33.

Richardson E, Pearce J, Mitchell R, Day P, Kingham S. 2010. The association between green space and cause-specific mortality in urban new zealand: An ecological analysis of green space utility. BMC public health 10:240.

Richardson EA, Pearce J, Mitchell R, Kingham SJPh. 2013. Role of physical activity in the relationship between urban green space and health. 127:318-324.

Robitzsch A, Grund S, Henke TJRpv. 2016. Miceadds: Some additional multiple imputation functions, especially for mice. 1:7-8.

Rothman KJ, Greenland S, Lash TL. 2008. Modern epidemiology:Lippincott Williams & Wilkins.

Su JG, Jerrett M, de Nazelle A, Wolch J. 2011. Does exposure to air pollution in urban parks have socioeconomic, racial or ethnic gradients? Environmental Research 111:319-328.

Su JG, Dadvand P, Nieuwenhuijsen MJ, Bartoll X, Jerrett MJEi. 2019. Associations of green space metrics with health and behavior outcomes at different buffer sizes and remote sensing sensor resolutions. 126:162-170.

Temam S, Varraso R, Pornet C, Sanchez M, Affret A, Jacquemin B, et al. 2017. Ability of ecological deprivation indices to measure social inequalities in a french cohort. BMC public health 17:956.

The United States Geological Survey. 2017. Landsat surface reflectance derived spectral indices. 3.

The United States Geological Survey. 2020. Landsat surface reflectance-derived spectral indices. Available: https://www.usgs.gov/land-resources/nli/landsat/landsat-normalized-difference-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con.

Therneau. 2000. Modeling survival data: Extending the cox model 2000 new york.

Therneau. 2019. Package 'survival'.

Tveite H. 2017. The qgis nnjoin plugin.

Twohig-Bennett, Caoimhe, and Andy Jones. "The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes." Environmental research 166 (2018): 628-637.

United Nations Population Division. 2014. World urbanization prospects: The 2014 revision, highlights. Department of economic and social affairs. 32.

Vienneau D, de Hoogh K, Faeh D, Kaufmann M, Wunderli JM, Röösli M, et al. 2017. More than clean air and tranquillity: Residential green is independently associated with decreasing mortality. Environment international 108:176-184.

White IR, Carlin JB. 2010. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Statistics in medicine 29:2920-2931.

Wu K-S, Huo X, Zhu G-HJSotte. 2008. Relationships between esophageal cancer and spatial environment factors by using geographic information system. 393:219-225.

List of Tables

Table 1. Baseline characteristics of the GAZEL cohort participants by their health status (has cancer or not) until 31 December 2016 (n= 19408).

Characteristics	Without cancer (n = 15333)	With cancer (n = 4075)	p-value
Sex (male)	10897 (71.1%)	3256 (79.9%)	< 0.001
Age at enrollment (years)	43.58 (3.52)	44.37 (3.36)	< 0.001
Follow-up duration (years)	27.46 (2.52)	20.54 (4.63)	< 0.001
Body mass index (kg/m^2)	25.48 (3.61)	25.79 (3.50)	< 0.001
Smoking status (smoker)	2698 (17.6%)	888 (21.8%)	< 0.001
Smoking intensity (pack-years)	0.15 (0.99)	0.21 (1.42)	< 0.001
Passive smoking (no)	7427 (48.4%)	1947 (47.8%)	0.454
Education			0.045
9-11 years of education	11271 (73.5%)	3058 (75.0%)	
12-13 years of education	1165 (7.6%)	253 (6.2%)	
14-15 years of education	887 (5.8%)	228 (5.6%)	
Other secondary diploma	1655 (10.8%)	440 (10.8%)	
Other diploma	355 (2.3%)	96 (2.4%)	

Characteristics	Without cancer (n = 15333)	With cancer (n = 4075)	p-value
Family status (in couple)	13176 (85.9%)	3572 (87.7%)	0.004
Socio-occupational Status			0.011
Executive	2541 (16.6%)	634 (15.6%)	
Supervisor	9055 (59.1%)	2361 (58.0%)	
Worker	3718 (24.3%)	1078 (26.5%)	
Fruit or vegetable use			0.065
Never or less than once a week	113 (0.7%)	40 (1.0%)	
Once or twice a week	1090 (7.1%)	303 (7.4%)	
More than twice a week but not every day	3524 (23.0%)	992 (24.3%)	
Every day or almost	10606 (69.2%)	2740 (67.2%)	
Alcohol consumption			< 0.001
Abstinent	1724 (11.2%)	390 (9.6%)	
Light drinker	6781 (44.2%)	1678 (41.2%)	
Moderate drinker	3457 (22.5%)	943 (23.1%)	
Heavy drinker	1776 (11.6%)	632 (15.5%)	
Unclear pattern	1595 (10.4%)	432 (10.6%)	
Population density			0.516
Urban	5716 (37.3%)	1492 (36.7%)	
Semi-urban	5214 (34.1%)	1375 (33.8%)	

Characteristics	Without cancer (n = 15333)	With cancer (n = 4075)	p-value
Rural	4376 (28.6%)	1199 (29.5%)	
Deprivation index class			0.872
First tertile	5135 (33.5%)	1377 (33.8%)	
Second tertile	5079 (33.1%)	1355 (33.3%)	
Third tertile	5118 (33.4%)	1343 (33.0%)	
Occupational exposure to carcinogens			< 0.001
None	8508 (56.0%)	2060 (51.0%)	
One	895 (5.9%)	227 (5.6%)	
Two	710 (4.7%)	230 (5.7%)	
Three or more	5090 (33.5%)	1522 (37.7%)	
$PM_{2.5}$ exposure ($\mu g/m^3$)	33.83 (5.74)	33.59 (5.91)	0.023
Distance to the nearest major road (km)	6.42 (1.58)	6.42 (1.59)	0.633
NDVI at 100m buffer (unitless)	0.46 (0.17)	0.47 (0.17)	0.056
NDVI at 300m buffer (unitless)	0.49 (0.16)	0.50 (0.16)	0.167
NDVI at 500m buffer (unitless)	0.51 (0.16)	0.51 (0.16)	0.092
NDVI at 1000m buffer (unitless)	0.53 (0.15)	0.53 (0.15)	0.265
Distance to forests(km)*	1.64 (2.01)	1.54 (1.90)	0.016
Distance to agriculture lands (km)*	0.92 (1.55)	0.85 (1.48)	0.073
Distance to urban greenspace (km)*	4.05 (4.96)	4.22 (5.17)	0.106

Summary statistics for continuous numeric variables are as mean (standard deviation) and for categorized variables are as number (percent with one decimal).

*: Calculated based on all participants regardless of area of residence.

Table 2. Correlation matrix (Pearson's/ Spearman's correlation coefficients) of the greenspace (NDVIs at different buffers), proximity to different green spaces (agricultural lands, urban green spaces, and forests), $PM_{2.5}$, and proximity to roads for participants at baseline.

	$PM_{2.5}$	**Proximity to road**	**NDVI (100m)**	**NDVI (300m)**	**NDVI (500m)**	**NDVI (1000m)**	**Distance to forests**	**Distance to agricultural lands**	**Distance to urban green spaces**
$PM_{2.5}$	1								
Proximity to road	0.24	1							
NDVI (100m)	-0.55	-0.2	1						
NDVI (300m)	-0.55	-0.19	0.92	1					
NDVI (500m)	-0.55	-0.18	0.88	0.96	1				
NDVI (1000m)	-0.54	-0.16	0.82	0.91	0.95	1			
Proximity to forests	-0.38	-0.11	0.48	0.53	0.55	0.59	1		
Proximity to agricultural lands	-0.42	-0.11	0.49	0.55	0.58	0.6	0.55	1	
Proximity to urban green spaces	0.34	0.13	-0.31	-0.34	-0.34	-0.35	-0.21	-0.27	1

$PM_{2.5}$ (particulate matters with the aerodynamic diameter less than 2.5 μm) as $\mu g/m^3$; NDVI (normalized difference in vegetation index) is unitless and measured based on reflectance from Landsat satellites data; proximity to agricultural lands, urban green space
, and forests were defined as a (1/ distance) calculated from CORINE land cover maps of 1990, 2000, 2006, and 2012; proximity to roads defined as an inverse of the direct distance of geocoded locations to the nearest road.
*: Correlations for NDVIs are based on Pearson's correlation. Correlations for distances are based on Spearman's rank test.

Table 3. Results of extended Cox models (hazard ratio and 95% confidence intervals) for the association between exposure to greenspace (with 10 years delay of 10-years moving average of exposure term; per an IQR increase for NDVI, and an IQR increase of proximity to different green spaces).

Cancer site	N cancer	NDVI at 100m	Proximity to agricultural lands	Proximity to forests	Proximity to urban green spaces
All	4075	1.075 (1.016, 1.137)	1.026 (1.002, 1.052)	1.038 (1.004, 1.074)	1.002 (0.969, 1.035)
Prostate	1401	1.148 (1.042, 1.265)	1.032 (0.987, 1.078)	1.020 (0.960, 1.083)	0.990 (0.938, 1.044)
Breast	378	0.824 (0.687, 0.989)	0.997 (0.935, 1.064)	0.959 (0.880, 1.044)	1.039 (0.910, 1.186)
Colorectal	367	0.991 (0.823, 1.194)	0.977 (0.905, 1.054)	1.049 (0.936, 1.176)	0.945 (0.853, 1.047)
Bladder	268	1.242 (0.998, 1.545)	0.980 (0.885, 1.085)	1.098 (0.941, 1.283)	0.980 (0.870, 1.105)
Lung	237	0.870 (0.691, 1.095)	1.010 (0.913, 1.118)	0.997 (0.867, 1.147)	1.015 (0.882, 1.169)
Skin	156	1.095 (0.813, 1.475)	1.102 (0.973, 1.248)	1.179 (0.982, 1.416)	1.189 (0.950, 1.489)

All hazard ratios (HRs) are estimated per IQR (NDVI100m: 0.216; proximity to agricultural lands:0.90; proximity to urban green space: 4.07; proximity to forests: 1.73). All models are sex-stratified and adjusted for smoking status, smoking intensity (pack per year), passive smoking, alcohol drinking, socio-occupational status, marital status, body mass index, vegetable consumption, education, occupational exposure to carcinogens, age at enrolment, 10 years cumulative exposure to air pollution ($PM_{2.5}$), distance to major roads, population density, and deprivation (based on French deprivation index at 2009).

Table 4. Results of time-dependent Cox models (hazard ratio and 95% confidence intervals) for the association between exposure to greenspace (with 10 years lag of 10-years moving average of exposure term) based on urban; semi-urban and rural area of residence for all-sites and site-specific cancers (prostate, breast, colorectal).

Cancer site	Urbanity	N cancer	NDVI at 100m	Proximity to agricultural lands	Proximity to forests	Proximity to urban green space
All	Urban	1492	1.069 (0.966, 1.182)			1.001 (0.890, 1.126)
	Semi-urban	1375	1.109 (1.015, 1.212)			1.010 (0.968, 1.054)
	Rural	1199	1.074 (0.982, 1.175)	1.007 (0.952, 1.065)	1.029 (0.956, 1.107)	
Prostate	Urban	469	1.179 (0.980, 1.419)			0.895 (0.754, 1.062)
	Semi-urban	485	1.179 (1.011, 1.374)			1.018 (0.951, 1.090)
	Rural	446	1.095 (0.941, 1.273)	1.020 (0.923, 1.127)	1.003 (0.888, 1.134)	
Breast	Urban	191	0.878 (0.668, 1.154)			0.938 (0.656, 1.341)
	Semi-urban	111	0.849 (0.631, 1.142)			1.046 (0.869, 1.259)
	Rural	74	0.926 (0.671, 1.277)	0.885 (0.781, 1.002)	0.898 (0.744, 1.084)	
Colorectal	Urban	143	0.883 (0.624, 1.250)			1.846 (0.962, 3.541)
	Semi-urban	121	0.938 (0.692, 1.272)			0.983 (0.863, 1.119)
	Rural	102	0.989 (0.743, 1.315)	0.994 (0.829, 1.191)	1.027 (0.807, 1.307)	

All hazard ratios (HRs) are estimated per an IQR for NDVIs (NDVI100m:0.216; NDVI300m:0.226; NDVI500m:0.223; NDVI1000m:0.216), and proximity to agricultural lands, urban green space, and forests. All models are sex-stratified and adjusted for smoking status, smoking intensity (pack per year), passive smoking, alcohol drinking, socio-occupational status, marital status, body mass index, vegetable consumption, education, occupational exposure to carcinogens, age at enrolment, 10 years cumulative exposure to air pollution ($PM_{2.5}$), distance to major roads, and deprivation (based on French deprivation index at 2009).

*: Nine subjects had no data on urban/ rural classification, therefore removed from the analyses.

NA: not applicable

Figure 1. Study area and geographical distribution of the GAZEL cohort participants at enrolment (1989).

Figure 2. Flow-chart of final population selection for analyses (n=19408).

Figure S1. Schematic describing exposure allocation pattern according to different datasets of the GAZEL cohort participants.

Figure 1. Study area and geographical distribution of the GAZEL cohort participants at enrolment (1989).

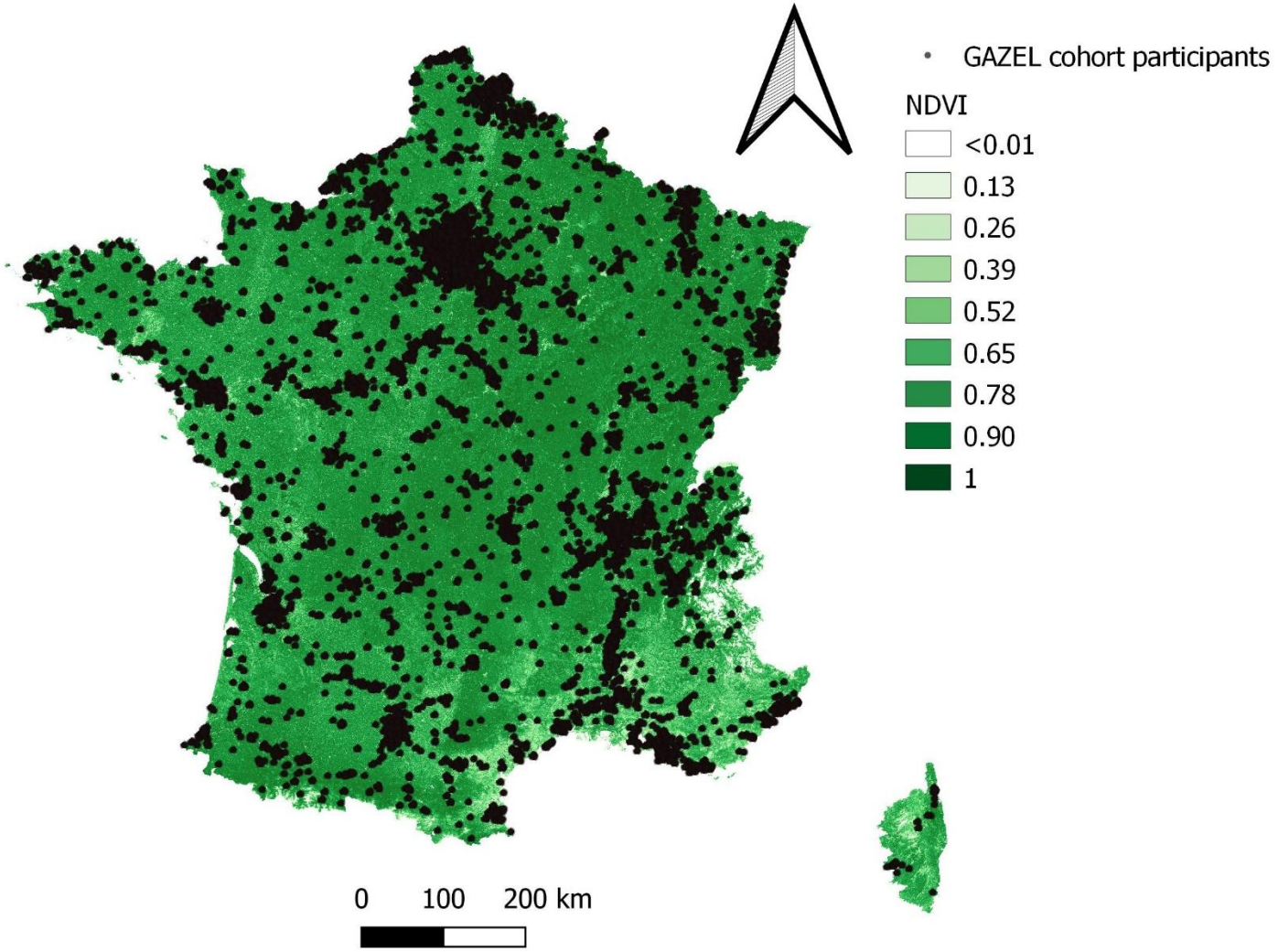
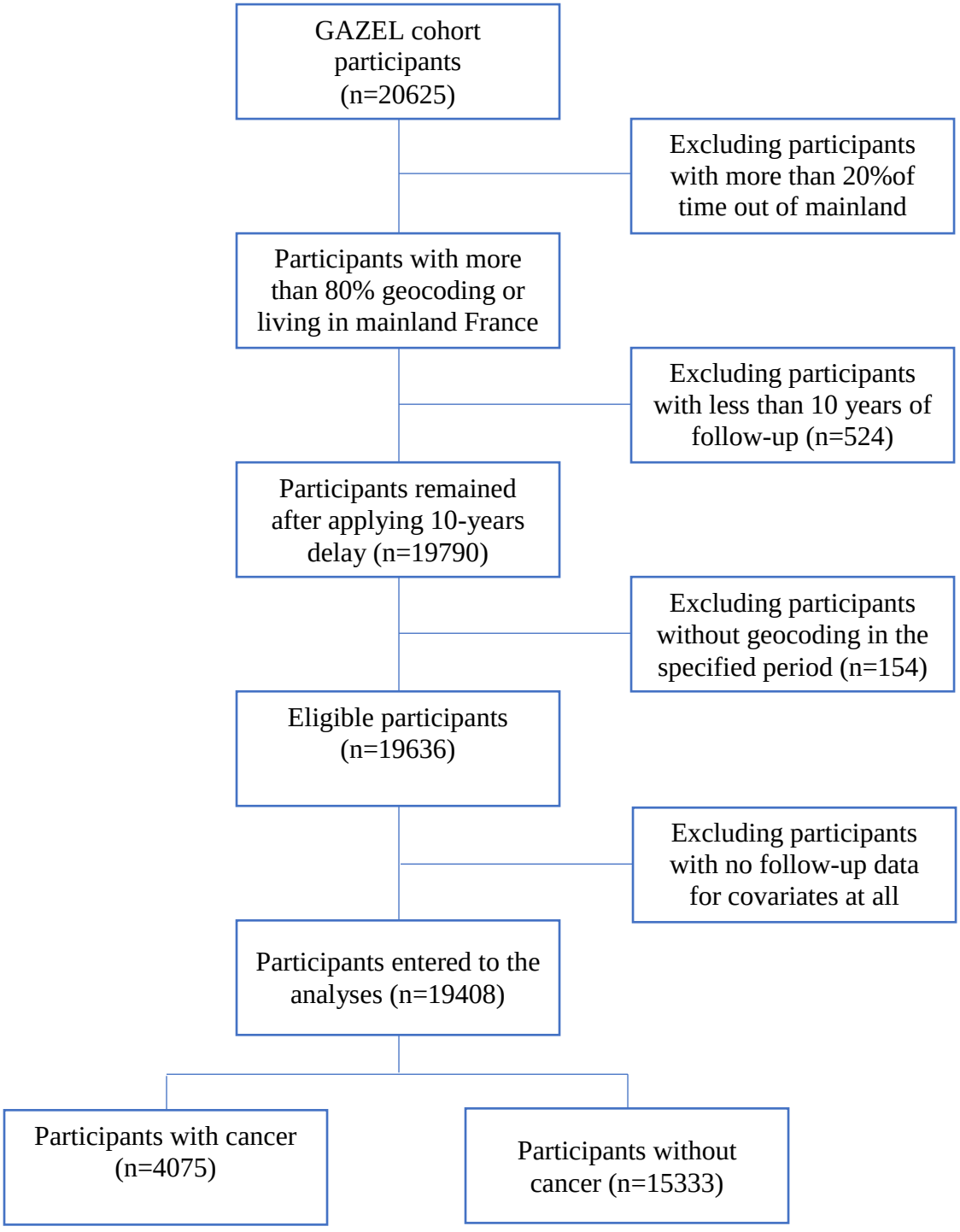


Figure 2. Flow-chart of final population selection for analyses (n=19408).



Authors Contribution:
Mohammad Javad Zare Sakhvidi: Conceptualization; Methodology; Data curation; Software;Formal analysis; Visualization; Writing – original draft; Writing – review & editing

Jun Yang: Conceptualization; Methodology; Writing – review & editing

Jack Siemiatycki: Methodology; Writing – review & editing
Payam Dadvand: Methodology; Writing – review & editing

Kees de Hoogh:Conceptualization; Methodology; Writing – review & editing; Resources

Danielle Vienneau:Conceptualization; Methodology; Writing – review & editing; Resources
Marcel Goldberg:Conceptualization; Methodology; Writing – review & editing; Investigation; Resources

Marie Zins:Conceptualization; Methodology; Writing – review & editing; Investigation; Resources

Emeline Lequy: Conceptualization; Methodology; Writing – review & editing

Bénédicte Jacquemin: Funding acquisition; Supervision; Project administration; Conceptualization; Methodology; Writing – review & editing

Declaration of interests

☒ 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.

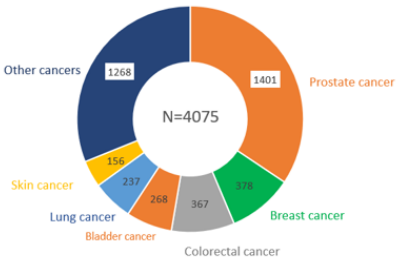
☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Graphical abstract

Participants and Exposure

GAZEL cohort data (n=19408)
NDVI and Land-cover data

Outcomes



4075 cases of cancer after 27 years of follow-up

Results

Potentially protective for:
Breast cancer

Increased risk of:
All-sites cancers

Heterogenous findings for:
Bladder cancer
Prostate cancer

The type of associations depends on type of greenspace and cancer site.

Highlights

- Evidences on the association between greenspace exposure and cancer is inconsistent.
- The association depends on the cancer site and greenspace type**.**
- Greenspace seems to have a protective role for the breast cancer**.**
- Greenness was linked to increased risk of all-site cancers.