Spatial and Spatio-temporal Epidemiology 9 (2014) 37–45

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

Case Study

A comparison between self-reported and GIS-based proxies of residential exposure to environmental pollution in a case–control study on lung cancer

M. Cordioli [a,b,*], A. Ranzi [b], A. Freni Sterrantino [b,c], L. Erspamer [b], G. Razzini [d], U. Ferrari [d], M.G. Gatti [e], K. Bonora [e], F. Artioli [d], C.A. Goldoni [e], P. Lauriola [b]

[a] *University of Parma, Department of Bio-Sciences, Parco Area delle Scienze 11/A, 43124 Parma, Italy*
[b] *Environmental Health Reference Centre, Regional Agency for Environmental Prevention of Emilia-Romagna, Via Begarelli 13, 41121 Modena, Italy*
[c] *University of Bologna, Department of Statistical Sciences, Via Belle Arti 41, Bologna, Italy*
[d] *Clinical Trials Office, Cancer Unit of the Carpi General Hospital, Via Guido Molinari 2, Carpi, Modena, Italy*
[e] *Department of Public Health, Unit of Epidemiology, Local Health Unit of Modena, Strada Martiniana 21, Baggiovara, 41126 Modena, Italy*

ARTICLE INFO

Article history:
Received 31 August 2013
Revised 3 March 2014
Accepted 25 April 2014
Available online 9 May 2014

Keywords:
Exposure assessment
GIS
Lung cancer
Questionnaire
Self-reported exposure
Environmental epidemiology

ABSTRACT

In epidemiological studies both questionnaire results and GIS modeling have been used to assess exposure to environmental risk factors. Nevertheless, few studies have used both these techniques to evaluate the degree of agreement between different exposure assessment methodologies.

As part of a case–control study on lung cancer, we present a comparison between self-reported and GIS-derived proxies of residential exposure to environmental pollution.

649 subjects were asked to fill out a questionnaire and give information about residential history and perceived exposure. Using GIS, for each residence we evaluated land use patterns, proximity to major roads and exposure to industrial pollution. We then compared the GIS exposure-index values among groups created on the basis of questionnaire responses.

Our results showed a relatively high agreement between the two methods. Although none of these methods is the "exposure gold standard", understanding similarities, weaknesses and strengths of each method is essential to strengthen epidemiological evidence.
© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Lung cancer has been associated with exposure to various environmental risk factors (Alberg et al., 2007; Field and Withers, 2012). Although less significant than smoke and other environmental agents (e.g., occupational exposure to carcinogens, radon, asbestos, etc.) outdoor air pollution is considered a possible risk factor in the etiology of this pathology (Pope III et al., 2011; Raaschou-Nielsen et al., 2011; Turner et al., 2011). Many studies suggest relative risks up to 1.5 for high versus low estimates of exposure to air pollution (Boffetta and Nyberg, 2003).

Various methodologies have been applied in studies on lung cancer to assess human exposure to possible environmental risk factors. These include: (i) self reported exposure (Chan-Yeung et al., 2003; Hosgood III et al., 2010; Hosseini et al., 2009); (ii) comparison between subjects living in urban versus rural areas (Curwen et al., 1954;

* Corresponding author at: Environmental Health Reference Centre, Regional Agency for Environmental Prevention of Emilia-Romagna, Via Begarelli 13, 41121 Modena, Italy. Tel.: +39 059 4226453; fax: +39 059 241312.

E-mail addresses: mik.cordioli@gmail.com (M. Cordioli), aranzi@arpa.emr.it (A. Ranzi), anna.freni@unibo.it (A. Freni Sterrantino), lerspamer@arpa.emr.it (L. Erspamer), g.razzini@ausl.mo.it (G. Razzini), u.ferrari@ausl.mo.it (U. Ferrari), gi.gatti@ausl.mo.it (M.G. Gatti), k.bonora@ausl.mo.it (K. Bonora), f.artioli@ausl.mo.it (F. Artioli), c.goldoni@ausl.mo.it (C.A. Goldoni), plauriola@arpa.emr.it (P. Lauriola).

http://dx.doi.org/10.1016/j.sste.2014.04.004
1877-5845/© 2014 Elsevier Ltd. All rights reserved.

Katsouyanni and Pershagen, 1997); (iii) proximity analysis (Garcia-Perez et al., 2009; Simonsen et al., 2010); and (iv) more sophisticated estimates of air-pollution levels (Nyberg et al., 2000; Hystad et al., 2012; Pope III et al., 2011; Raaschou-Nielsen et al., 2011; Turner et al., 2011). Among others, Geographic Information Systems (GIS) are being used with increasing frequency in environmental epidemiology (Nuckols et al., 2004).

One of the main advantages of using GIS versus collecting subjective measures of exposure is the possibility to evaluate personal exposure in large populations. Both self-reported and GIS-derived exposure may be affected by errors, the first mainly due to subjective perception of risks, the latter because of poor data quality. Although none of these methods can be considered the "gold standard" for exposure assessment to atmospheric pollution (Forastiere and Galassi, 2005), comparing different methods is of great interest because it allows the identification of the weakness and strengths of each method and the development of integrated exposure indicators.

Few studies have investigated the relationship between self-reported and modeled exposure to environmental pollution (Cesaroni et al., 2008; Gunier et al., 2006; Heinrich et al., 2005; Migliore et al., 2009). Published studies have focused on exposure to traffic-related pollution with regard to respiratory symptoms like asthma, cough, bronchitis or rhinitis, while none of these studies have referred to lung cancer.

In the present study, we evaluated the association between self-reported exposure to a variety of environmental risk factors and GIS-derived proxies of exposure in a case–control study on lung cancer. In view of the long latency period of this pathology, we devoted particular attention to reconstructing each subject's exposure history. The goals of this analysis were to evaluate the potential of GIS data in exposure assessment compared with self-reported information, and to highlight the importance of using multiple exposure assessment methods in epidemiological studies where "true" measures of personal exposure are not available.

2. Materials and methods

2.1. Study design and population

Data on cancer incidence for the Province of Modena (Northern Italy) in the years 2000–2005 showed a possible cluster for lung cancer in the District of Mirandola, where the Standardized Incidence Ratio (SIR) for males reached the value of 1.26 (CI 95%: 1.13–1.40) (Pirani et al., 2007). A prospective case–control study (the *IDEALE* project) was carried out to investigate the association between environmental risk factors and lung cancer in an area comprising the 9 municipalities belonging to the Mirandola Health District.

A total of 649 subjects were enrolled (case:control ratio of 1:4). Cases were defined as incident events of lung cancer in the period 2009–2010 and controls were coupled on the basis of sex and age. A summary of the main characteristics of the population enrolled is given Table 1.

Table 1
Summary of the main characteristics of the enrolled population.

	n	%
Number of subjects		
Total	649	100
Cases	130	20
Controls	519	80
Sex		
Males	504	78
Females	145	22
Age (years)		
<50	33	5
50–70	262	40
>70	354	55
Smoking habits		
Smoker	110	17
Ex-smoker	351	54
Non-smoker	188	29
Education		
None/Primary school	362	56
Junior high school	148	23
High school	119	18
Degree	20	3

Subjects were interviewed face to face, using a questionnaire designed to collect personal data and information about lifestyle, active and passive smoking habits, food and alcohol consumption, health status, residential and occupational history.

In particular, participants were asked to give information about exposure to environmental pollution at each address of residence since year 1980. The questions, which derived from standardized questionnaire formats (Erspamer et al., 2007; Goldoni et al., 2003; Migliore et al., 2009; SIDRIA, 1997;), were the following:

(1) The zone of residence is predominantly: rural/residential/industrial
(2) The street of residence is: busy/quiet
(3) Do the majority of windows look out directly onto busy roads: yes/no
(4) Are there crossroads or traffic lights within 100 m of the house: yes/no
(5) Do you find dust on windowsills: always or frequently/sometimes/never

2.2. Reconstruction of residential history

Each address reported in the questionnaire was geocoded (coordinate system: UTM32, datum ED50) through record-linkage by street name and street number to the Regional Database (RDB) of the Emilia Romagna Region. Since some of the municipalities in the study area were not included in this database, some addresses were directly geocoded using a global positioning system (GPS); those remaining were geocoded using free web services (Google Maps and Microsoft Bing).

2.3. Exposure assessment

In the whole study area there was only one fixed air-pollution monitoring station, so these data were not usable to differentiate the geographic variability of exposure in enrolled subjects. Thus, for each residence we evaluated exposure to possible environmental risk factors by adopting a geographical approach based on GIS (ArcGIS® v.9.3, post-elaboration with Stata SE® v.12 and R v.2.13.1). We focused on three possible risk factors: land use in the neighborhood of the residence (Eberhardt and Pamuk, 2004; Mitchell and Popham, 2007), road traffic (Cook et al., 2011; Kim et al., 2008; Nuvolone et al., 2011), and industrial pollution (Hendryx and Fedorko, 2011; Luo et al., 2011; Pronk et al., 2013).

We also collected available information about land use, roads, and industrial emissions for the municipalities within 5 km of the study area, in order to correctly characterize exposure for residences located near the study area boundaries. Past residences of enrolled subjects located outside the study area and over 5 km from it were excluded from the exposure assessment.

2.3.1. Road traffic

Exposure to traffic was defined through road proximity analysis (Bayer-Oglesby et al., 2006; Cook et al., 2011). Since no homogeneous and sufficiently detailed information on traffic flows were available for the area under study, we used the Regional road network cartography (RER, 2011), defining the importance of each road based on the functional classification used by the administration. According to this classification, roads were defined as:

- "major" (i.e., highways and important roads used by national or regional traffic);
- "secondary" (i.e., streets connecting city centers or used by cross-city traffic);
- "minor" (i.e., small roads used to reach a specific address or location).

Since there was only one major road in the entire study area, we finally grouped roads into two classes, i.e., (i) "major and secondary roads" (*magsec*) and (ii) "minor roads" (*minor*).

For each residence we computed the following variables (Fig. 1-A): typology of the nearest road; minimum distance from a *magsec* road (variable *mindist_magsec*); sum of the length of all streets and *magsec* streets inside a buffer within a radius of 100 m and 200 m around the residence (*all_100/200* and *magsec_100/200*).

To test the association between the two exposure assessment methods, we compared the distributions of GIS-derived exposure indexes between the groups created on the basis of self-reported proxies of exposure. Since GIS variables showed non-normal distributions, we used non parametric statistical tests: the Wilcoxon rank-sum test (one tailed, $\alpha = 0.05$) was used to test the difference between two groups (questions 2, 3 and 4) while the Cuzik's test for trend ($\alpha = 0.05$) was used to test the presence of a trend in exposure within three groups (question 5).

2.3.2. Land use in the neighborhood

The prevalent land use inside three concentric buffers within a radius of 250, 500 and 1000 m around each residence was characterized by means of an intersection with the Regional land use cartography (RER, 2011) for the years 1976, 1994, 2003 and 2008 (Fig. 1B). Each map was considered as representative respectively of the periods: <1985, 1986–1999, 2000–2006, 2007–2011.

The original Corine Land Cover categories (EEA, 2000) were divided into 4 new groups: agricultural/natural green areas (*aver*), dump/construction sites (*esdi*), industrial/commercial (*indu*) and urban fabric/urban green areas (*urpar*). The coverage by each land use group was then computed as a percentage (%) of the total area of each buffer so as to define the prevalent land use (i.e., at least 20% more area than all the other three groups) within each buffer for each period.

We then assessed the degree of association between answers to question (1) and the prevalent land use within each buffer by means of the Chi-square test or Fisher's exact test ($\alpha = 0.05$).

To compare data with the questionnaire responses, we chose the land use map which corresponded to the last year of each period of residence. As a sensitivity analysis the central year of each residence was also considered.

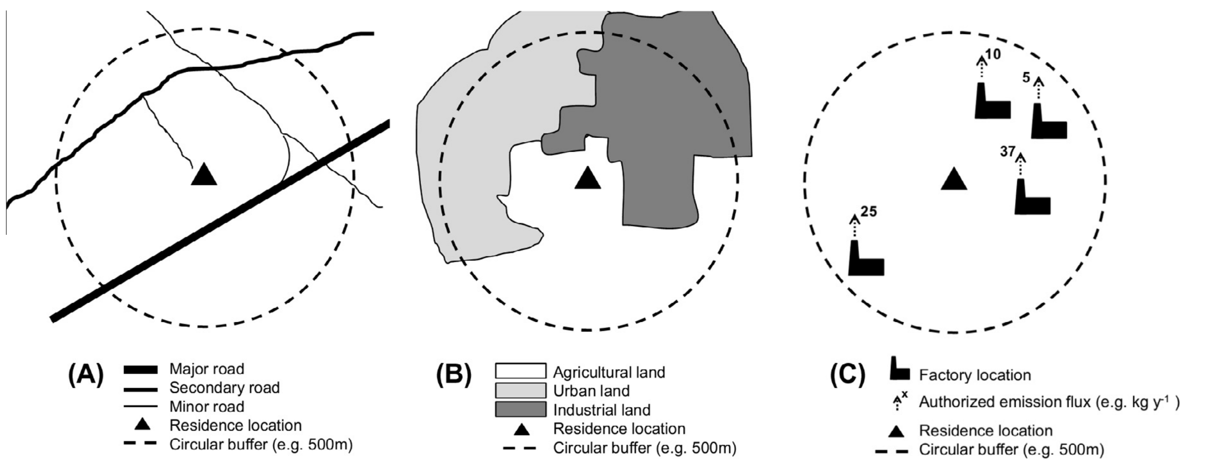

Fig. 1. Schematic representation of GIS exposure assessment: (A) length of roads by typology inside a buffer around the residence, (B) percentage of buffer area covered by different land uses, (C) tons of pollution emitted inside a buffer.

2.3.3. Industrial pollution

Exposure to industrial pollution was characterized, evaluating the amount of pollutants emitted into the atmosphere by factories and commercial activities around the place of residence (Agarwal et al., 2010; Hendryx and Fedorko, 2011; Luo et al., 2011; Pronk et al., 2013; Willis et al., 2010) (Fig. 1C).

We collected information on authorized emission fluxes (i.e., based on Italian regulations DPR 203/88 and D.lgs 152/2006) of total suspended particles (TSP) and volatile organic compounds (VOC) from the Regional Environmental Protection Agency. Additional information about the start and end date for each activity was obtained from the local Chamber of Commerce. The addresses of industrial plants were geocoded using the methodology described above for residences.

For each residence, and for each year, we computed the total flux of TSP and VOC (tons year^{-1}) emitted into the atmosphere inside three buffers within a radius of 100, 500 and 1000 m. We considered emission fluxes for each industrial plant as constant over time, but we used available information on the start and end date of both residences and industrial activities to modulate the inter-annual variability inside the buffers. For each buffer we then computed the average emission rate over the period of residence (*ptsov100/500/1000)*.

We evaluated the association between the answers to questions (1) and (5) and the GIS-derived exposure (categorized as "presence" or "absence" of emissions in the buffer) using the Chi-square test or Fisher's exact test ($\alpha = 0.05$). We evaluated the presence of a trend in total emission fluxes within the groups defined by answer to question (5) using Cuzik's test for trend ($\alpha = 0.05$).

3. Results

3.1. Residential address geocoding

The 649 subjects interviewed reported a total of 838 residential addresses. The population can be considered quite stable: 73% of subjects have never changed residence since 1980, 25% moved once, and 3% had more than 2 residences. 37 residences fell outside the study area and the 5 km buffer zone.

A total of 801 residential addresses inside the study area were geocoded, 60% using the Regional database (RDB), 9% using GPS and 30% using free web services. Fig. 2 shows the study area and the location of the residences.

To assess the degree of comparability between the results of the three geocoding methods, for one subset of residences we computed the Euclidean distance between the coordinates derived from Google and those derived from the RDB database ($n = 129$) or the GPS ($n = 50$). The median error for the Google-RDB and Google-GPS comparison was respectively 28 m and 35 m, while the 90th percentile was respectively 559 m and 196 m. In some limited cases different geocoding methods yielded very

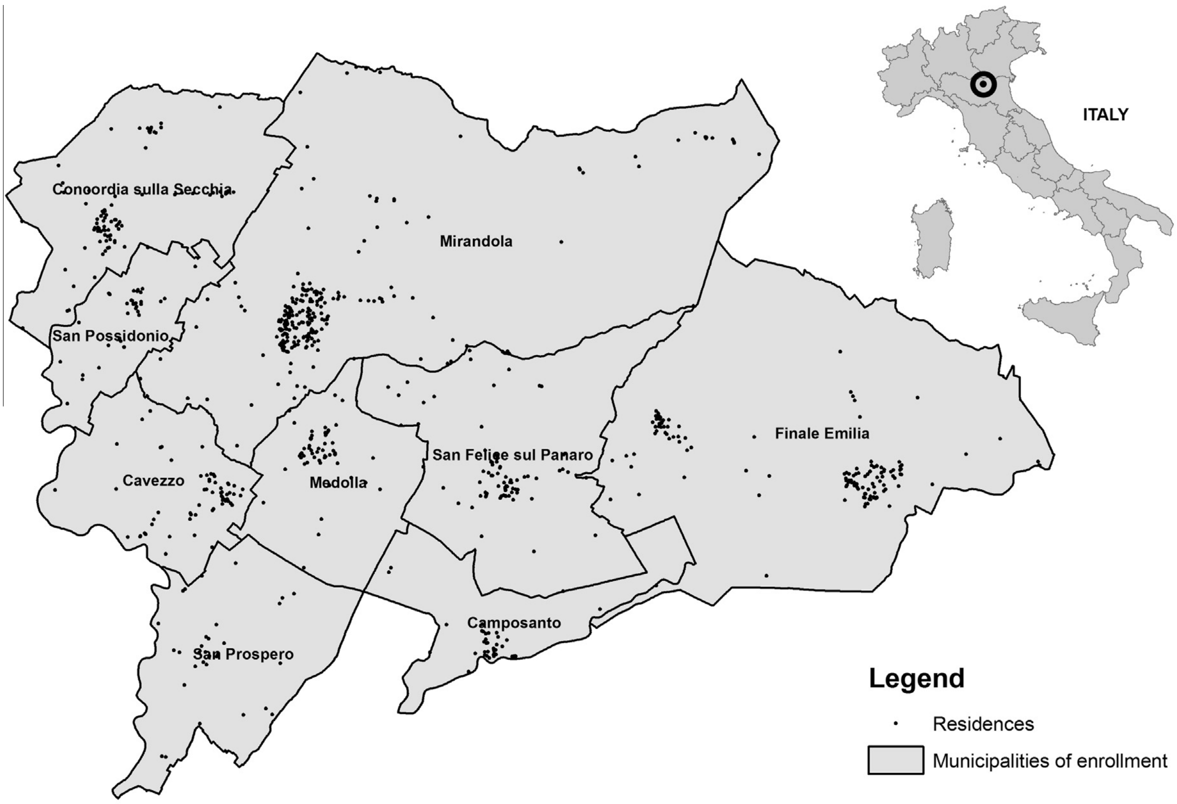

Fig. 2. Representation of the study area and distribution of residences in the 9 municipalities of enrolment.

different coordinates (up to 2683 m distance). Fig. S3 in Supplementary Information represents the cumulative distribution of calculated distances.

3.2. Comparison between questionnaire and GIS exposure assessment

Table S1 in Supplementary Information reports frequencies for the answers to the questionnaire, grouped by the educational level and disease status. Questionnaire answers were not differently distributed between cases and control subjects (Piro et al., 2008), while answers to questions (1) and (4) were differently distributed within educational levels (i.e., people with higher education more frequently reported residences in "urban" and "industrial" areas and the presence of crossroads).

3.2.1. Land use in the neighborhood

Table 2 shows frequencies of GIS-derived prevalent land use inside three buffers grouped by answer to question (1).

Almost all the residences were classified as *aver* or *urpar*, regardless of the buffer chosen.

For all considered buffers, frequencies of prevalent land use show very different distributions within the groups defined by answer to question (1) (Chi-square test highly significant). Bold values in Table 2 represent the percentage of matching between self reported and GIS-derived classification of land use. The maximum level of matching for "residential" and "industrial" residences, i.e., 86% and 67% respectively, was observed for the 250 m buffer. An increase in the buffer radius yielded a decrease in the percentage of matching for these land use classes. On the other hand, when the largest buffer was used nearly all the residences (i.e., 99%) declared in the questionnaire as "rural" were correctly classified in the *aver* group.

Table S2 in Supplementary Information shows the same calculation performed using land-use maps for the intermediate year of each residence. Overall, the level of matching decreased, especially for the "residential" and "urban" respondents.

Residences of more educated subjects yielded slightly higher frequencies of *urpar* and *indu* prevalent land use compared with less educated subjects (e.g., for *urpar* 81% and 63% respectively for degree and none/primary school), in accordance with self-reported data (Table S1).

3.2.2. Road traffic

Table 3 shows the results for exposure assessment to traffic. The residences were at an average distance of 182 m from *magsec* roads in the area, with an average of 97 m and 328 m of *magsec* roads respectively within the 100 m and 200 m buffer.

All exposure variables showed highly significant ($p < 0.01$) or significant ($p < 0.05$) differences between the groups created on the basis of self-reported proxies of exposure, the sole exception being the total length of all streets inside buffers (i.e., *all_100* and *all_200*) for question (2). Only for 26% of the residences reported to be on a "busy road" was the nearest street identified with GIS a *magsec* road, whereas 92% of residences reported to be on a "quiet road" were closer to a *minor* road.

For question (5), about the presence of dust on windowsills, all variables showed a highly significant trend (Cuzick's test, $p < 0.01$) between groups. Compared with residences in which dust is frequently found, residences without reported dust are on average at twice the distance from a *magsec* road. For these residences also the total length of *magsec* roads inside the 100 m buffer is on average 75% lower.

3.2.3. Industrial pollution

We identified a total of 897 industrial and commercial activities with authorized TSP and VOC emissions. Pollution sources were geocoded using the RDB (78%) and Google (22%). Evaluation of exposure to industrial pollution showed that only 5% of residences had authorized TSP and VOC emission sources within a distance of 100 m, while this percentage increased to 34% and 65% respectively for the 500 m and 1,000 m buffer (Table 4, classified as "yes"). For the exposed residences, the median levels of TSP plus VOC emissions inside the three buffers was respectively 0.15, 0.17 and 2.4 t year^{-1} (Table 4). All residences with *indu* prevalent land use inside the 500 m buffer had emission sources within the same distance, while this was true only for 39% of *urpar* and 24% of *aver* residences.

Table 2
Results for the characterization of the area of residence, considering land-use maps for the last year of each residence. For each buffer, the number represents the fraction of residences in each prevalent land-use category (i.e., row percentages). Bold values represent the land-use categories we consider matching with each response to the questionnaire. *p*-Values refer to the Chi-squared test of association.

Zone of residence [a]	Prevalent land use inside buffer								
	250 m			500 m			1000 m		
	aver [b]	urpar [c]	indu [d]	aver	urpar	indu	aver	urpar	indu
All ($n = 801$)	31%	65%	3%	44%	54%	2%	74%	26%	0%
Rural ($n = 222$)	**82%**	17%	1%	**91%**	9%	0%	**99%**	1%	0%
Residential ($n = 561$)	12%	**86%**	2%	25%	**74%**	1%	63%	**37%**	0%
Industrial ($n = 18$)	17%	17%	**67%**	33%	22%	**44%**	100%	0%	**0%**
p-Value Chi.	$p < 0.01$		$p < 0.01$	$p < 0.01$					

[a] Responses to question (1). Total number of subjects may differ due to to incomplete responses.
[b] Agricultural/natural green areas.
[c] Urban fabric/urban green areas.
[d] Industrial/commercial areas.

Table 3
Indexes of exposure to road traffic. Mean (standard deviation) are shown for each group. p-Values refer to the one-tailed Wilcoxon test for the difference between groups (questions 2, 3, and 4) or to the Cuzick's test for trend (question 5).

Question	Answer		Traffic exposure indexes				
		n[a]	all_100[b] (m)	magsec_100[c] (m)	all_200[d] (m)	magsec_200[e] (m)	mindist_magsec[f] (m)
None		801	588 (234)	97 (106)	1936 (806)	328 (262)	182 (241)
Street type (2)	Busy	308	586 (224)	133 (111)	1933 (770)	401 (254)	124 (186)
	Quiet	493	589 (240)	75 (97)	1938 (828)	282 (257)	219 (264)
		p-Value Wilc.	$p = 0.54$	$p < 0.01$	$p = 0.52$	$p < 0.01$	$p < 0.01$
Windows (3)	Yes	264	616 (220)	132 (109)	2050 (780)	410 (256)	124 (186)
	No	526	573 (241)	80 (100)	1874 (816)	287 (256)	211 (261)
		p-Value Wilc.	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$
Crossings (4)	Yes	387	650 (222)	107 (112)	2189 (741)	373 (275)	141 (158)
	No	411	529 (231)	88 (100)	1698 (793)	285 (242)	222 (295)
		p-Value Wilc.	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.01$	$p < 0.05$
Dust (5)	Always/frequently	106	615 (196)	153 (120)	2061 (684)	450 (282)	103 (145)
	Sometimes	127	631 (212)	93 (111)	2199 (750)	359 (288)	173 (258)
	Never	540	571 (241)	87 (100)	1856 (822)	298 (246)	201 (252)
		p-Value trend	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$	$p < 0.01$

[a] Total number of subjects may differ due to to incomplete responses.
[b] Length of all streets inside the 100 m buffer.
[c] Length of major and secondary streets inside the 100 m buffer.
[d] Length of all streets inside the 200 m buffer.
[e] Length of major and secondary streets inside the 200 m buffer.
[f] Distance to the nearest major or secondary street.

Table 4
Results for exposure to industrial pollution. Numbers represent row percentages of residences for each group. For exposed residence, the median flux of TSP and VOC in tons year^{-1} is shown in parenthesis. p-Values refer to the Chi-squared test of association or to the Cuzick's test for trend for pollution fluxes.

Question	Answer		Industrial activities exposure indices					
			ptsov100[b]		ptsov500[c]		ptsov1000[d]	
		n[a]	No	Yes	No	Yes	No	Yes
None		801	95%	5% (0.2)	66%	34% (0.2)	35%	65% (2.4)
Zone (1)	Rural	222	95%	5% (0.1)	80%	20% (0.2)	59%	41% (0.9)
	Residential	561	97%	3% (0.0)	63%	37% (0.1)	26%	74% (2.0)
	Industrial	18	50%	50% (5.3)	6%	94% (12.1)	0%	100% (24.9)
	p-Value Chi/F		$p < 0.01$		$p < 0.01$		$p < 0.01$	
Dust (5)	Always/frequently	106	92%	8% (0.2)	67%	33% (0.5)	25%	75% (7.9)
	Sometimes	127	94%	6% (4.1)	62%	38% (0.1)	30%	70% (1.6)
	Never	540	96%	4% (0.1)	67%	33% (0.2)	37%	63% (1.5)
	p-Value Chi/F		$p = 0.29$		$p = 0.58$		$p < 0.05$	
	p-Value Trend		$p = 0.49$		$p = 0.04$		$p = 0.01$	

[a] Total number of subjects may differ due to to incomplete responses.
[b] PTS and SOV emission inside the 100 m buffer.
[c] PTS and SOV emission inside the 500 m buffer.
[d] PTS and SOV emission inside the 1000 m buffer.

The proportion of residences with TSP and VOC emissions is lower for houses declared as "rural" and higher for houses declared as "residential" and "industrial". The Pearson Chi-squared test confirmed a different distribution of the exposure indices in the three groups defined by answers to question (1), with a stronger concordance on "industrial" area when buffers increased. Consistently, the highest median level of emission for each buffer radius was found for residences reported as "industrial" (Table 4).

No relation was found between the presence of emission sources and the presence of dust on windowsills (question (5)), except for the 1000 m buffer, where the proportion of exposed residences was slightly higher in the "always/frequently" group. On the other hand, the level of emission around exposed residences showed a statistically significant positive trend (Cuzik's test, $p < 0.05$) towards increasing levels of reported dust for the 500 m and 1000 m buffer.

4. Discussion

This work presented a comparison between self-reported and GIS-derived proxies of exposure to environmental pollution from a case–control study on lung cancer. To our knowledge, this is the first study on lung cancer to have used both questionnaire and GIS to derive proxies of exposure to multiple environmental risk factors. Previous studies applied questionnaires to derive residential history and GIS to assess exposure (Bellander et al., 2001; Hystad et al., 2012; Nyberg et al., 2000) but no data on self-perceived exposure to environmental risks was collected (environmental tobacco smoke and occupational exposure excluded).

Some previously published works have addressed the problem of validation of self-reported exposure, but their focus was normally on traffic-related pollution only. Heinrich et al. (2005) reported a relatively low level of agreement between self-reported exposure to traffic and a combination of air-pollution measurement and GIS modeling, with subjective assessment overestimating the modeled exposure, especially in the symptomatic group. Gunier et al. (2006) reported a good correspondence between self-reported traffic exposure and GIS-based traffic-density measurements, while there was no evidence of association between self-reported exposure and levels of Polycyclic Aromatic Hydrocarbons (PAH) in the urine. Cesaroni et al. (2008) analyzed the association between various indices of exposure to traffic and respiratory health in Rome, highlighting a statistically significant association between self-reported exposure and modeled proxies of exposure. Migliore et al. (2009) found no evidence of bias in a validation analysis where they compared self-reported traffic intensity with measured traffic fluxes. Recently, Birk et al. (2011) reported a fair agreement between self-reported noise annoyance and GIS-modeled noise levels in Munich.

Overall, our results showed moderate to good agreements between the information from the questionnaire and exposure indices computed using spatial analyses.

The comparison between GIS-derived prevalent land use around the residence and subjective characterization of the neighborhood gave good results. In particular, residences reported as "rural" were surrounded by a large area of prevalently agricultural land (i.e., the 1000 m buffer), while the perception of living in a "residential" or "industrial" context was better related to the characteristics of the immediate surroundings (i.e., the 250 m buffer). Thus, if land use characteristics in a larger area around the residence are believed to be important in determining local air quality, exposure assessment should be more effectively based on GIS rather than on self-reported data.

We replicated the analysis reported in Table 2, which refers to prevalent land use in the last year of each residence, also using land-use maps for the intermediate year between each residence (Table S2 in Supplementary Information). Interestingly, the matching quota was reduced for "residential" and "industrial" responses (i.e., respectively 77% and 50% for the 250 m buffer). In fact, there was a remarkable modification in land use inside the study area during the last years, with an increase in *urpar* and *indu* land-use types and a corresponding decrease in the *aver* category. For example, between 1976 and 2008, 96% of residences experienced a reduction in the *aver* land use inside the 500 m buffer, 91% an increase in the *urpar* land use, while 60% experienced an increase in the *indu* land use. The higher level of matching obtained using more recent maps (Table 2 versus Table S2) may be due to the fact that people more easily remember the characteristics of the residence in the latter years. This is of interest, since lung cancer has a long induction–latency period and past exposures may be more relevant.

Almost all the GIS-derived indices of traffic exposure were coherently associated with self-reported proxies. Residences located on "busy" roads, with windows exposed to traffic or where crossroads were reported in the neighborhood, had more *magsec* streets inside the buffers and a minor distance to *magsec* roads. Interestingly, all indices of exposure showed a trend towards the categories of self-reported dust on windowsills. Some evidence of a correlation between the characteristics of indoor dust and outdoor road dust from traffic resuspension has been previously reported in the literature (Kuo et al., 2012; Sbihi et al., 2013; Tong and Lam, 2000). Nevertheless, our assessment was based on self-reported qualitative data and was insufficient to draw conclusions in this sense.

Finally, residences self-reported as "industrial" proved to be more exposed to industrial TSP and VOC emissions, while no association was found between the presence of dust on windowsills and the presence of industrial sources in the surroundings. Nevertheless, the analyses of quantitative data on emission fluxes suggest a possible association between increasing levels of dust on windowsills and the quantity of TSP and VOC emitted from industries inside the 500 m and 1000 m buffer.

It is difficult to identify a "gold standard" between self-reported information and GIS-derived models (Forastiere and Galassi, 2005). Both methodologies suffer from possible errors. The collection of both exposure and disease data from questionnaires may be prone to so-called recall bias (Coughlin, 1990; Piro et al., 2008) and interviewer bias (Wynder, 1994), and this could lead to errors in risk estimations (Kuehni et al., 2006; Rugbjerg et al., 2011).

In this sense, the use of GIS is a promising development since it yields more objective exposure readings. However, GIS modeling is also subject to errors and limitations (Krieger, 2003). In our study we evaluated the performance of different geocoding systems, with results that are in agreement with data on positional accuracy from other published studies (Duncan et al., 2011; Whitsel et al., 2006). Shifting the position of residences by some hundred meters may lead to errors in the definition of proximity-based indices of exposure (Zandbergen, 2007; Zanderbergen and Green, 2007).

Moreover, some streets classified as *minor* may have been characterized by intense traffic due to the presence of specific poles of attraction or traffic lights, while incomplete data on some industrial emissions and their evolution in time may have led to overestimating or underestimating exposure.

Another shortcoming in GIS use may derive from the absence of information about past exposure. In our study, we had good quality information about past land use in the area and relatively complete records about industrial emission. Using such data, we were able to model the evolution of exposure in time. This would not have been possible if we had used only self-reported information. On the other hand, no data were available on the evolution of the road network over recent years.

The direct comparison between subjective and objective assessment presents certain limitations. To compute the prevalent land use inside a buffer, we used an *a priori* cut-off (i.e., 20% more surface then all the other categories), and it is unlikely that this procedure would have been used in self-assessment. Furthermore, the nearest street identified through GIS was not necessarily the street to which people referred in answering question (2), since the actual entrance to the house may have been in another street.

Finally, one shortcoming of our approach may be represented by the fact that both self-reported and GIS-derived exposure variables may exhibit a certain degree of spatial autocorrelation. Spatial autocorrelation occurs when characteristics at proximal locations appear to be correlated, so that near things are more similar than distant things. The spatial nature of our data may thus bias the results of the statistical tests used (Dale and Fortin, 2009).

5. Conclusions

The use of GIS in epidemiological research is an interesting challenge for improvement in the quality of analyses and reliability of findings. Nevertheless, it is important to provide evidence of the validity of these new tools with respect to "classical" epidemiological methods. The results of our work showed moderate to good agreements between GIS-derived proxies of exposure to environmental risks and self-reported evaluation from a questionnaire. We were able to highlight some peculiarity of self-reported exposure, e.g., a tendency in interviewed people to refer to more recent land use characteristics and to consider the presence of industrial areas in the immediate vicinity of the home address. Moreover, we identified some agreements between self-reported presence of dust on home window sills and proxies of exposure to road traffic.

Both the methods employed in this study have their specific weaknesses and we could not define which was the best one. Nevertheless, finding a good agreement between different methodologies of exposure assessment is essential to strengthen epidemiological evidence. Even if the information about environmental risk factors in the area were in certain cases incomplete or approximate, the use of GIS represented a powerful tool in characterizing the exposure of residents. When good quality data are available, the use of objective measures of exposure instead of self-reported data is encouraged. GIS and spatial analysis enable considerations about the temporal and spatial dimension of environmental exposure, which subjective evaluation cannot manage/handle. In particular, the use of these spatial proxies of exposure is essential in cases of non-availability of self-reported exposure, not uncommon in epidemiological studies involving a larger number of enrolled subjects.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The IDEALE project was funded by Fondazione Cassa di Risparmio di Mirandola and by Associazione Malati Oncologici (AMO). The Authors wish to thank Chiara Leporati from the Carpi-Mirandola Hospital for her meticulous work in selecting and contacting subjects to be enrolled in the study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.sste.2014.04.004. These data include Google maps of the most important areas described in this article.

References

Agarwal N, Banternghansa C, Bui LTM. Toxic exposure in America: Estimating fetal and infant health outcomes from 14 years of TRI reporting. J Health Econ 2010;29:557–74. http://dx.doi.org/10.1016/j.jhealeco.2010.04.002.

Alberg AJ, Ford JG, Samet JM. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 2007;132:29S–55S.

Bayer-Oglesby L, Schindler C, Hazenkamp-von Arx ME, Braun-Fahrlander C, Keidel D, Rapp R, et al. Living near main streets and respiratory symptoms in adults the swiss cohort study on air pollution and lung diseases in adults. Am J Epidemiol 2006;164:1190–8. doi:10.1093/aje/kwj338.

Bellander T, Berglind N, Gustavsson P, Jonson T, Nyberg F, Pershagen G, et al. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. Environ Health Perspect 2001;109:633–9.

Birk M, Ivina O, Von Klot S, Babisch W, Heinrich J. Road traffic noise: self-reported noise annoyance versus GIS modelled road traffic noise exposure. J Environ Monit 2011;13:3237–45. http://dx.doi.org/10.1039/c1em10347d.

Boffetta P, Nyberg F. Contribution of environmental factors to cancer risk. Br. Med. Bull. 2003;68:71–94. http://dx.doi.org/10.1093/bmb/ldg023.

Cesaroni G, Badaloni C, Porta D, Forastiere F, Perucci CA. Comparison between various indices of exposure to traffic-related air pollution and their impact on respiratory health in adults. Occup Environ Med 2008;65:683–90. http://dx.doi.org/10.1136/oem.2007.037846.

Chan-Yeung M, Koo LC-M, Ho JC-M, Tsang KW-T, Chau W-S, Chiu S-W, Ip MS-M, Lam W-K. Risk factors associated with lung cancer in Hong Kong. Lung Cancer 2003;40:131–40. http://dx.doi.org/10.1016/S0169-5002(03)00036-9.

Cook AG, Jbm A, Pereira G, Jardine A, Weinstein P. Use of a total traffic count metric to investigate the impact of roadways on asthma severity: a case–control study. Environ Health 2011;10:52. http://dx.doi.org/10.1186/1476-069X-10-52.

Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol 1990;43:87–91. http://dx.doi.org/10.1016/0895-4356(90)90060-3.

Curwen MP, Kennaway EL, Kennaway NM. The incidence of cancer of the lung and larynx in urban and rural districts. Br J Cancer 1954;8:181–98.

Dale MRT, Fortin M-J. Spatial autocorrelation and statistical tests: some solutions. J Agric Biol Environ Statist 2009;14:188–206. http://dx.doi.org/10.1198/jabes.2009.0012.

Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL. Evaluation of the positional difference between two common geocoding methods. Geospat Health 2011;5:265–73.

Eberhardt M, Pamuk E. The importance of place of residence: examining health in rural and nonrural areas. Am J Public Health 2004;94:1682–6.

EEA. CORINE land cover technical guide - Addendum 2000. Tecnical report No 40. Copenhagen; 2000.

Erspamer L, Ranzi A, Silvi G, Rubbi L, Lauriola P. Case–control study on the long-term effects of urban-industrial pollution: the Ravenna experience. Epidemiol Prev 2007;21:28–33 [In Italian].

Field RW, Withers BL. Occupational and environmental causes of lung cancer. Clin Chest Med 2012;33:681–703. http://dx.doi.org/10.1016/j.ccm.2012.07.001.

Forastiere F, Galassi C. Self report and GIS based modelling as indicators of air pollution exposure: is there a gold standard? Occup Environ Med 2005;62:508–9. http://dx.doi.org/10.1136/oem.2005.019901.

Garcia-Perez J, Pollàn M, Boldo E, Perèz-Gòmez B, Aragonès N, Lope V, et al. Mortality due to lung, laryngeal and bladder cancer in towns

lying in the vicinity of combustion installations. Sci Total Environ 2009;407:2593–602. http://dx.doi.org/10.1016/j.scitotenv.2008.12.062.
Goldoni C, Gambini M, Vaccari C, Soli A, Ranzi A, De Girolamo G, et al. The role of epidemiology applied to a local area based on a study caried out in the community of Fiorano Modenese (ceramic district). Epidemiol Prev 2003;27:310–5 [In Italian].
Gunier RB, Reynolds P, Hurley SE, Yerabati S, Hertz A, Strickland P, et al. Estimating exposure to polycyclic aromatic hydrocarbons: a comparison of survey, biological monitoring, and geographic information system – based methods. Cancer Epidemiol Biomarkers Prev 2006;15:1376–81. http://dx.doi.org/10.1158/1055-9965.EPI-05-0799.
Heinrich J, Gehring U, Cyrys J, Brauer M, Hoek G, Fischer P, et al. Exposure to traffic related air pollutants: self reported traffic intensity versus GIS modelled exposure. Occup Environ Med 2005;62:517–24. http://dx.doi.org/10.1136/oem.2004.016766.
Hendryx M, Fedorko E. The relationship between toxics release inventory discharges and mortality rates in rural and urban areas of the united states. J Rural Health 2011;27:358–66. http://dx.doi.org/10.1111/j.1748-0361.2011.00367.x.
Hosgood III HD, Boffetta P, Lee YA, Mclaughlin J, Seow A, Duell EJ, et al. In-Home Coal and Wood Use and Lung Cancer Risk: A Pooled Analysis of the International Lung Cancer Consortium. Environ Health Perspect 2010;118:1743–7. http://dx.doi.org/10.1289/ehp.1002217.
Hosseini M, Naghan PA, Karimi S, Seyedalinaghi S, Bahadori M, Khodadad K, et al. Environmental risk factors for lung cancer in Iran: a case–control study. Int J Epidemiol 2009;38:989–96. http://dx.doi.org/10.1093/ije/dyp218.
Hystad P, Demers PA, Johnson KC, Brook J, VanDonkelaar A, Lamsal L, et al. Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case–control study. Environ Health 2012;11. http://dx.doi.org/10.1186/1476-069X-11-22.
Katsouyanni K, Pershagen G. Ambient air pollution exposure and cancer. Cancer Causes Control 1997;8:284–91. http://dx.doi.org/10.1023/A:1018492818416.
Kim JJ, Huen K, Adams S, Smorodinsky S, Hoats A, Malig B, et al. Residential traffic and children's respiratory health. Environ Health Perspect 2008;116:1274–9. http://dx.doi.org/10.1289/ehp.10735.
Krieger N. Place, space and health: GIS and epidemiology. Epidemiology 2003;14 [DOI: 0.1097/01.ede.0000071473.69307.8a].
Kuehni CE, Strippoli MF, Zwahlen M, Silverman M. Association between reported exposure to road traffic and respiratory symptoms in children: evidence of bias. Int J Epidemiol 2006;35:779–86. http://dx.doi.org/10.1093/ije/dyl022.
Kuo C-Y, Wang J-Y, Liu W-T, Lin P-Y, Tsai C-T, Cheng M-T. Evaluation of the vehicle contributions of metals to indoor environments. J Eposure Sci Environ Epidemiol 2012;22:489–95. http://dx.doi.org/10.1038/jes.2012.55.
Luo J, Hendryx M, Ducatman A. Association between six environmental chemicals and lung cancer incidence in the United States. J Environ Public Health 2011;2011:463701. http://dx.doi.org/10.1155/2011/463701.
Migliore E, Berti G, Galassi C, Pearce N, Forastiere F, Calabrese R, et al. Respiratory symptoms in children living near busy roads and their relationship to vehicular traffic: results of an Italian multicenter study (SIDRIA 2). Environ Health 2009;8:27. http://dx.doi.org/10.1186/1476-069X-8-27.
Mitchell R, Popham F. Greenspace, urbanity and health: relationships in England. J Epidemiol Community Health 2007;61:681–3. http://dx.doi.org/10.1136/jech.2006.053553.
Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. Environ Health Perspect 2004;112:1007–15. http://dx.doi.org/10.1289/ehp.6738.
Nuvolone D, Maggiore R, Maio S, Fresco R, Baldacci S, Carrozzi L, et al. Geographical information system and environmental epidemiology: a cross-sectional spatial analysis of the effects of traffic-related air pollution on population respiratory health. Environ Health 2011;10:12. http://dx.doi.org/10.1186/1476-069X-10-12.
Nyberg F, Gustavsson P, Järup L, Bellander T, Berglind N, Jakobsson R, et al. Urban air pollution and lung cancer in Stockholm. Epidemiology 2000;11:487–95.
Pirani M, Rashid I, Cirilli C, Federico M. Analisi geografica dei tumori diagnosticati in provincia di Modena negli anni 2000–2005. Modena; 2007 [In Italian].
Piro FN, Madsen C, Næss , Nafstad P, Claussen B. A comparison of self reported air pollution problems and GIS-modeled levels of air pollution in people with and without chronic diseases. Environ Health 2008;7:9. http://dx.doi.org/10.1186/1476-069X-7-9.
Pope III CA, Burnett RT, Turner MC, Cohen A, Krewski D, Jerrett M, et al. Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships. Environ Health Perspect 2011;119:1616–21. http://dx.doi.org/10.1289/ehp.1103639.
Pronk A, Nuckols JR, De Roos AJ, Airola M, Colt LS, Cerhan JR, et al. Residential proximity to industrial combustion facilities and risk of non-Hodgkin lymphoma: a case–control study. Environ Health 2013;12:20. doi:10.1186/1476-069X-12-2.
Raaschou-Nielsen O, Andersen ZJ, Hvidberg M, Jensen SS, Ketzel M, Sørensen M, et al. Lung cancer incidence and long-term exposure to air pollution from traffic. Environ Health Perspect 2011;119:860–5. http://dx.doi.org/10.1289/ehp.1002353.
RER (Regione Emilia Romagna). Servizio Sviluppo dell'Amministrazione Digitale e Sistemi Informativi Geografici; 2011. Available at: http://geoportale.regione.emilia-romagna.it/ [date last accessed: 15/07/2013].
Rugbjerg K, Harris MA, Shen H, Marion SA, Tsui JKC, Teschke K. Pesticide exposure and risk of Parkinson's disease – a population-based case–control study evaluating the potential for recall bias. Scand J Work Environ Health 2011;37:427–36. http://dx.doi.org/10.5271/sjweh.3142.
Sbihi H, Brook JR, Allen RW, Curran JH, Dell S, Mandhane P, et al. A new exposure metric for traffic-related air pollution? An analysis of determinants of hopanes in settled indoor house dust. Environ Health 2013;12:48. http://dx.doi.org/10.1186/1476-069X-12-48.
SIDRIA. Asthma and respiratory symptoms in 6–7 yr old Italian children: gender, latitude, urbanization. Eur Respir J 1997;10:1780–6. http://dx.doi.org/10.1183/09031936.97.10081780.
Simonsen N, Scribner R, Su LJ, Williams D, Luckett B, Yang T, et al. Environmental exposure to emissions from petrochemical sites and lung cancer: the lower mississippi interagency cancer study. J Environ Public Health 2010;2010. http://dx.doi.org/10.1155/2010/759645. ID: 759645.
Tong STY, Lam KC. Home sweet home? A case study of household dust contamination in Hong Kong. Sci. Total Environ. 2000;256:115–23. http://dx.doi.org/10.1016/S0048-9697(00)00471-X.
Turner MC, Krewski D, Pope CA, Chen Y, Gapstur SM, Thun MJ. Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. Am J Respir Clin Care Med 2011;184:1374–81. http://dx.doi.org/10.1164/rccm.201106-1011OC.
Whitsel EA, Quibrera PM, Smith RL, Catellier DJ, Liao D, Henley AC, et al. Accuracy of commercial geocoding: assessment and implications. Epidemiol Perspect Innovat 2006;3:8. http://dx.doi.org/10.1186/1742-5573-3-8.
Willis AW, Evanoff BA, Lian M, Galarza A, Wegrzyn A, Schootman M, et al. Metal emissions and urban incident parkinson disease: a community health study of medicare beneficiaries by using geographic information systems. Am J Epidemiol 2010;172:1357–63. http://dx.doi.org/10.1093/aje/kwq303.
Wynder EL. Investigator bias and interviewer bias: The problem of reporting systematic error in epidemiology. J Clin Epidemiol 1994;47:825–7. http://dx.doi.org/10.1016/0895-4356(94)90184-8.
Zandbergen PA. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. BMC Public Health 2007;7:37. http://dx.doi.org/10.1186/1471-2458-7-37.
Zanderbergen PA, Green JW. Error and bias in determining exposure potential of children at school locations using proximity-based GIS techniques. Environ Health Perspect 2007;115:1363–70. http://dx.doi.org/10.1289/eph.9668.