



METHODOLOGY

Open Access

Gravity models to classify commuting vs. resident workers. An application to the analysis of residential risk in a contaminated area

Guido Signorino¹, Roberto Pasetto², Elisa Gatto^{1*}, Massimo Mucciardi¹, Marina La Rocca¹, Pierpaolo Mudu³

Abstract

Background: The analysis of risk for the population residing and/or working in contaminated areas raises the topic of commuting. In fact, especially in contaminated areas, commuting groups are likely to be subject to lower exposure than residents. Only very recently environmental epidemiology has started considering the role of commuting as a differential source of exposure in contaminated areas. In order to improve the categorization of groups, this paper applies a gravitational model to the analysis of residential risk for workers in the Gela petrochemical complex, which began life in the early 60s in the municipality of Gela (Sicily, Italy) and is the main source of industrial pollution in the local area.

Results: A logistic regression model is implemented to measure the capacity of Gela "central location" to attract commuting flows from other sites. Drawing from gravity models, the proposed methodology: a) defines the probability of finding commuters from municipalities outside Gela as a function of the origin's "economic mass" and of its distance from each destination; b) establishes "commuting thresholds" relative to the origin's mass. The analysis includes 367 out of the 390 Sicilian municipalities. Results are applied to define "commuters" and "residents" within the cohort of petrochemical workers. The study population is composed of 5,627 workers. Different categories of residence in Gela are compared calculating Mortality Rate Ratios for lung cancer through a Poisson regression model, controlling for age and calendar period. The mobility model correctly classifies almost 90% of observations. Its application to the mortality analysis confirms a major risk for lung cancer associated with residence in Gela.

Conclusions: Commuting is a critical aspect of the health-environment relationship in contaminated areas. The proposed methodology can be replicated to different contexts when residential information is lacking or unreliable; however, a careful consideration of the territorial characteristics ("insularity" and its impact on transportation time and costs, in our case) is suggested when specifying the area of application for the mobility analysis.

Introduction

Epidemiological studies have extensively considered mobility, intended as the change of individual residence, as a risk factor for health. Mobility accounts for accessibility to health services [1-3], and is widely used as a proxy for different times of exposure to environmental pollutants over an individual's lifetime. Furthermore, human mobility has been analyzed as one of the most important vehicles of infectious diseases [4-8]. More

generally, the relationship between migration and health is controversial; higher morbidity or mortality rates within in-migration areas, compared with territories showing stable populations, are detected by Boyle et al. [9,10] and Davey Smith et al. [11], in contrast with other evidence [12-15]. Brown and Leyland [16] show that small areas with lower mobility rates in Scotland have better health outcomes, and Martikainen et al. [17] discuss the importance of selective migration when studying the relationship between area socioeconomic characteristics and individual health profiles.

From a methodological perspective, Rogerson and Han [18] argue that mobility may hamper the detection of

* Correspondence: egatto@unime.it

¹Department of Economics, Statistics, Mathematics and Sociology, University of Messina, Via Tommaso Cannizzaro 278, Messina, Italy
Full list of author information is available at the end of the article

both geographical differences in disease diffusion and regional variability in disease risk, while Jacquez et al. [19,20] censure “the static world-view in which individuals are considered immobile, migration between populations does not occur, and in which background disease risks under the null hypothesis are assumed to be time-invariant and uniform through geographic space”, and propose representation methods of human mobility that account for space-time relations when studying residential epidemiology of long latency illnesses, such as cancer. In their study, “participants must have lived in the study area for at least the past 5 years and had no prior history of cancer”.

However, mobility implies not only that people change their residence, but also that (increasingly, with transport facility improvements) they commute on a daily basis, travelling from their residence to a “distant” workplace. From this perspective, commuting has been investigated as a factor that influences individuals’ exposure during travel time [21-26], as a source of stress [27] that increases cardio-vascular (CDV) risk [28-30] and travel accident risk [31,32], or as an opportunity to stimulate physical activity both for adult [33-37] and for young people [38-40]. Commuting is also relevant in the identification of different exposure levels to some specific pollutant for populations working within highly polluted areas [41]. This is especially true in the case of “contaminated areas”, i.e. areas characterised by the presence of polluting industrial activities; in these sites commuting groups are subject to different exposure periods than resident workers and share “occupational” exposure with their fellows, but have a lower “environmental” exposure than residents, as they live in non-contaminated areas during their non-working time.

Commuting is, then, an important topic in environmental epidemiology and may help to shed light on residential/occupational health risks for different groups. Nevertheless, even though commuting has been intensively studied from geographic (see, among others, [42-44]) and economic perspectives (see Rowendal and Nijkamp [45] for a review of this literature), to our knowledge, there have been no studies attempting to relate commuting and health on the basis of geo-economic models. Therefore, this paper represents an attempt to bridge territorial studies with public health.

With the aim of integrating geo-economic quantitative methodologies with epidemiological studies, this study offers a multidisciplinary perspective in order to improve the definition of groups of exposed populations in cases in which the reconstruction of individuals’ residential history is not possible (either due to missing or unreliable information).

Drawing from mobility gravity models, the paper proposes a quantitative method based on the use of a

logistic probability model, to define a possible valuable classification of a population of petrochemical workers into commuters and residents categories; on these grounds, we then test residential risk for lung cancer within a cohort of workers from the Gela petrochemical complex (in Sicily, Italy). More generally, the application of commuting analysis may be important for epidemiological investigations, and particularly for occupational epidemiology.

Background

In Italy, the region of Sicily (which is the biggest island in the Mediterranean) contains three areas that have been declared “at high risk of environmental crisis” (“risk areas”), located in three different portions of the territory, as shown in Figure 1.

Referring to the specific Italian legislation, “risk areas” include territories where the presence of large industrial plants has an important impact on the environment and may harm human health. The general criteria for the determination of the perimeters of risk areas are open to debate and must always be scrutinised. The formal recognition by law of the condition of risk in a particular area, follows specific risk assessment procedures. Sometimes risk areas are identified after the occurrence of particular events raising environmental concern, while in other cases, the recognition of environmental and health risk comes after a long “invisible” process of contamination or when significant epidemiological factors single out a population group living in a particular area [46].

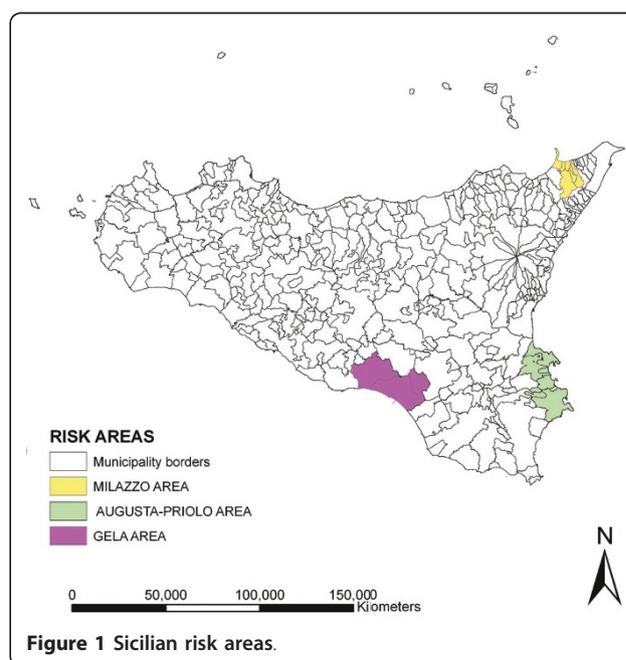


Figure 1 Sicilian risk areas.

In this study, we specifically concentrate on the site of Gela, a town of 77,000 inhabitants, located on the south-west coast of Sicily and characterised by relative isolation with respect to adjacent municipalities. The town underwent great changes in only a few years after 1956, when oil fields were discovered in its vicinity, both onshore and offshore; since the early 1960s, Gela has hosted a large oil refinery, together with a number of important chemical and petrochemical industries. The industrial area extends along a large portion of the territory in comparison to the size of the urban area.

In 1990, an extended territory comprising the municipalities of Gela, Butera and Niscemi (see Figure 2) was declared an “area at high risk of environmental crisis” (Law n. 349/1986) in consideration both of the serious accidents that may occur at the petrochemical complex and their possible consequences on the local population.

In 1998, a portion of Gela municipality, including the entire petrochemical complex and a vast offshore area, was ruled a “site of national concern for soil remediation” (Law n. 426/1998) due to heavy chemical soil pollution derived from the petrochemical complex.

Several studies report information about environmental contamination levels in the area of Gela; in particular, two specific studies on pine needles and road dust samples showed that the entire area of Gela is heavily affected by industrial and urban emissions of metals and metalloids [47,48]; additionally, data on Gela air quality revealed the presence of high levels of SO₂, PM₁₀, NO₂, O₃ and benzene [49], while, as regards the area of concern for soil remediation, underground water and soil are highly contaminated by heavy metals (arsenic, mercury, nickel, manganese, iron, lead, aluminium), hydrocarbons, BTEX (benzene, toluene, ethylbenzene and xylenes) and carcinogenic chlorinated aliphatic compounds [50]. Some of these water and soil pollutants

(arsenic, mercury, benzene, nickel) greatly exceed legal limits, thus representing a potential source of health risk for the resident population.

More generally, evidence of a risk of lung cancer, associated with residence near petrochemical plants has been suggested by several studies. A higher risk of lung cancer was observed in people residing in the most polluted areas of Teesside in the UK [51,52] and in case-control studies carried out in the U.S.A., Taiwan and Italy [53-55]. In these studies, exposure to petrochemical industries was indirectly evaluated using the distance of residences from the plants and duration of residence as proxies.

Usually, the contribution of the epidemiological investigation of occupational cohorts is related to the occupational risk evaluated in the production setting under study [56].

In the industrial site of Gela, the large petrochemical complex is the main source of environmental pollution. Some small-area mortality studies of the Gela resident population showed an excessive risk of lung cancer among men and women [57-59], showing also an excessive morbidity risk for acute and chronic diseases of the respiratory system among Gela residents [58,59].

More recent studies have shown that commuting may influence the result of small-area studies and should be carefully considered in the identification of the control population. Selecting areas with similar socio-economic characteristics and with no (or low) commuting to Gela, obtains higher Standardized Mortality Ratios than using neighbouring municipalities [60].

In this frame, the analysis of workers’ mortality and/or morbidity profiles can help to elucidate the environmental/residential risk, thus contributing to the description of the local epidemiological context. In fact, a more intense exposure should be assumed for workers residing within the contaminated area compared with commuters, giving rise to the expectation of a higher lung cancer mortality-rate among resident workers.

In general (in Italy as well as in many other countries), information from employment rosters may neither be used to reconstruct residential history, nor to define residence at the moment of enrolment. The latter in particular is true for several reasons: a) workers may be hired before the official change of residence takes place (and there is no need for firms to update their internal documents); b) electoral legislation encourages poll participation by subsidizing travel for people living away from their official place of residence; c) many people may have a strong preference to maintain residence in their birthplace in order to have closer contact with their original family.

In the case of the Gela petrochemical complex, for instance, the employment rosters show that there are

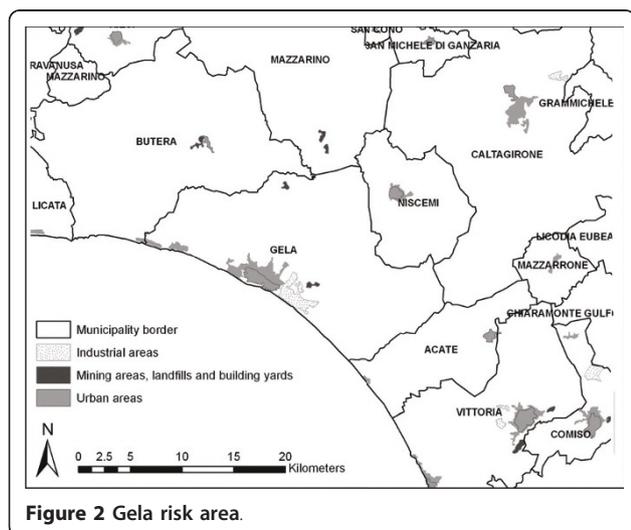


Figure 2 Gela risk area.

workers officially residing at a great distance from Gela, and that Gela's poor communication infrastructures (no motorway, no electrified railway) make it impossible for commuters to reach their workplace, even if they reside in the same region. As a result, residence information from the employment rosters is an unreliable criterion for a qualitative classification of workers as commuters vs. residents, while the simple use of place of birth as a classification criterion, overestimates the number of commuters. Alternative methodologies need to be developed.

Evidence about commuting among workers was initially inferred from several documents and successively confirmed by a qualitative sociological study [61].

Pasetto et al. [62] analyzed mortality within the cohort of petrochemical workers defining "residents" as all workers that were born in Gela or outside Sicily, and "commuters" all workers that were born in another municipality in the same region. This crude classification was due to the lack of a valid criterion to define commuters and raised the problem of the misclassification of residential categories [63,64].

In what follows, a methodology based on the application of mobility models is developed, in order to obtain a more accurate classification criterion to separate "commuting" and "resident" workers, and it is applied to the Gela petrochemical workers' cohort.

Methods

Commuting Probability

Mobility is a kind of spatial interaction whose analysis generally relies upon the application of gravity-type schemes [65,66] that consider the quantity of interactions between distant territories as a positive function of their economic force of attraction (or "economic mass"), and as a negative function of the distance that separates them. Economic mass can be expressed using any sound gross measure of economic activity or endowment (local GDP, population, labour force, natural resources, quantity and quality of services per inhabitant, etc.); in the applied literature [67], GDP, population and labour force are the variables most in use to proxy economic mass. Formally, gravitational models for spatial interaction are derived from the Newtonian framework according to the following general formula:

$$T_{ij} = A(i)B(j)/F(d_{ij}) \quad (1)$$

where T_{ij} is the flow of "interactions" between locations i and j , $A(i)$ and $B(j)$ are unspecified origin and destination weight functions, which may contain centre attributes relative to its economic mass, and $F(d_{ij})$ is defined as a *distance deterrence* (or *distance-decay*)

function, which accounts for the effect of distance d on T . The specification and combination of origin and destination functions may vary depending on the assumptions made about model parameters and variables. For an in-depth review of the theoretical aspects of gravity models, see Fotheringham and O'Kelly [68].

Gravity models and distance-decay functions have been applied to a wide range of research fields, from the analysis of social phenomena [69,70] to biology and environmental science [71-73], international trade [74,75] urban planning [76] and commuting [67,77,78]. In the field of health and health-care, gravity-type models have been used to build accessibility measures and study spatial accessibility to primary care [79], or to deepen the spatial-temporal dynamics of epidemics in inland towns and coastal cities in England and Wales [80].

As far as the distance-decay function specification is concerned, both the measure of spatial separation and the specific functional form of the relation have been debated [68,81]. For the first aspect, it is possible to use different measures: distance (defined in terms of physical, cultural, social or even religious dimensions), travel cost and travel time. These variables can be combined in various ways [67], but in most applications, physical distance is used as a proxy for both travel cost and travel time [68]. Concerning the formal specification of the distance-decay relation, there is a general consensus about the use of either power or exponential functions (equations (2) and (3)):

$$F(d_{ij}) = d_{ij}^{-\beta} \quad (2)$$

$$F(d_{ij}) = \exp(-\beta d_{ij}) \quad (3)$$

where the parameter β measures the effect of distance friction. Structural properties of the exponential function make it appropriate to model short distance interactions (intra-urban mobility), while power specification is more suitable for longer distance interactions such as migration flows [68].

Within an unconstrained setting, the complete model can be structured as follows for power or exponential decay functions, respectively:

$$T_{ij} = v_i^\mu w_j^\alpha d_{ij}^{-\beta} \quad (4)$$

$$T_{ij} = v_i^\mu w_j^\alpha \exp(-\beta d_{ij}) \quad (5)$$

From an empirical point of view, a widely used approach to investigate spatial interaction, is to calibrate the model using log-linear regression [81,68]; alternatively, probabilistic models based on Poisson regression [82-84] have been applied.

Our aim in implementing the mobility analysis, is to derive a methodological device that can be used to classify workers into “residents”/“commuters” categories in case the only reliable information is relative to the individuals’ birthplace. To this purpose, a gravity-type model is implemented to estimate the probability for a “central locality”, where a specific plant is established, to attract commuting flows from other distant places. Probability of commuting will then become the criterion to discriminate observations in the cohort study, according to the workers’ place of birth. Working on census data, we need to obtain probabilities from count values, so that a logistic regression that accounts for the origin’s economic mass and distance from destination is the most appropriate functional form for this kind of application of gravity-type models.

In our case, we are firstly interested in elucidating to what extent Gela increases commuting flows from “distant” localities in order to identify a critical distance (the “commuting threshold”), discriminating petrochemical workers according to their birthplace. In fact, assuming that Gela-born people employed at the petrochemical complex reside in Gela, workers born outside this municipality may have either moved their residence to Gela, or maintained their residence in a different site, regularly commuting to Gela. Having identified the commuting threshold, workers born outside that distance will be considered “moved to Gela”, while individuals born within this area will be classified as “commuters”.

Clearly, in the presence of a single destination, its characteristics are invariant in the model: the $B(j)$ component of model (1) disappears and mobility can be estimated as an unconstrained function of the origins’ economic mass ($A(i)$) and distance from Gela (d_{ij}).

Hence, having transformed Italian Statistics Bureau Census count data on inter-municipal commuting [85] into a dichotomous variable, with value 1 attributed to municipalities with positive commuting towards Gela, and value 0 assigned to municipalities with no commuting to Gela, a logistic regression model has been implemented in order to define the “probability of commuting”. The results of this analysis have been applied to classify as “commuting municipalities” all sites whose commuting probability is equal to or greater than 0.5 and “no commuting municipalities” all localities whose commuting probability is lower than 0.5.

The estimated model is then:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 dist + \beta_2 lf \quad (6)$$

where p_i and $1 - p_i$ are respectively the probability of observing a positive and a null flow of commuters to

Gela. According to the gravitational pattern, explanatory variables are, distance in kilometres ($dist$) of each municipality from Gela and a proxy for the origins’ “economic mass”, provided by labour force (lf); the latter is treated as a dummy variable on the basis of classification into quartiles (see Table 1).

In order to define the area that has to be considered when implementing the mobility analysis, we propose to identify the longest distance from which a commuter flow is originated and consider this as the array of an ideal circumference, whose area delimitates the boundary of the territorial analysis, excluding all municipalities whose distance from the destination pole is greater than this extension.

In order to limit the extension of our geographic analysis and to define the observations to be considered, we notice that the furthest locality from Gela where a positive commuting is found, is Trapani, at a distance of 248 km.; further, to reach Gela by car from the Italian peninsula, one has to cover a distance of approximately 500 km. at a cost of nearly 100 € per return journey, spending more than 7 hours (data from <http://Viamichelin.com>), while a one-way journey by train takes from 7 h38 m to 11 h41 m (information from Italian Railways official site). As a consequence, we have restricted the analysis to 367 Sicilian municipalities less than 248 km from Gela.

Mobility model results

The model is estimated with the maximum likelihood estimator using both the Newton-Raphson and Fisher scoring iterative algorithms; the two methods provide the same results confirming estimates robustness.

As illustrated in Table 2, the result of the Hosmer and Lemeshow Test and the Nagelkerke R^2 value (0.66) show a good fit of the estimated probabilities with respect to the empirical ones. The individual significance of coefficients, as measured by the Wald test, is very high, confirming the reliability of the estimated model. Moreover, the signs of coefficients are coherent with the theoretical background.

On the basis of the conventional probability cut-off of 0.5, we assume that no commuting originates from municipalities with a probability value lower than or equal to 0.5, so that, for each “economic mass” category, a threshold distance for commuting can be fixed in correspondence to the 0.5 estimated probability. The model

Table 1 Labour force classification

lf ₁	Very low (<than 853)
lf ₂	Low (between 853 and 1650)
lf ₃	Medium (between 1650 and 3830)
lf ₄	High (>than 3830) (reference variable)

Table 2 Commuting probability estimation results

Dependent variable: 1 = at least one commuter, 0 = otherwise

	β	Wald	Df	p-value	Exp(β)	95% C.I. for Exp(β)	
						Lower	Upper
<i>Dist</i>	-.040 (.005)	73.291	1	.000	.961	.952	.970
<i>lf*</i>		44.126	3	.000			
<i>lf(1)</i>	-2.713 (.502)	29.257	1	.000	.066	.025	.177
<i>lf(2)</i>	-2.691 (.474)	32.278	1	.000	.068	.027	.172
<i>lf(3)</i>	-2.331 (.469)	24.713	1	.000	.097	.039	.244
<i>Constant</i>	6.212 (.724)	73.664	1	.000	498.926		
LR test	243**		4	.000			
Nagelkerke R ²	.659						
Hosmer and Lemeshow Test	9.709		8	.286			

Standard errors are in parentheses.

*Reference category = High (>than 3830 units).

** Initial -2LL = 477, Final -2LL = 234.

produces four “distance thresholds” relative to the different economic mass consistency of the origin municipalities; these thresholds, as reported in Table 3, define, for each of the envisioned municipality categories, that distance from Gela beyond which commuting is not likely to be observed.

In accordance with the assumptions of the gravitational model, a greater dimension of the origin municipalities implies higher distance thresholds.

By comparing estimated and observed values, the model correctly classifies nearly 90% of observations (see Table 4).

Table 4 illustrates the model classification decision rule:

As a consequence, we suggest classifying as “commuters” all workers that were born in municipalities lying within the commuting threshold estimated for the mass category, and as “moved to Gela” all those workers who were born outside the threshold.

Results

The cohort study

Results of the mobility model were applied to the cohort of Gela petrochemical workers. The following categories of “presumed residence” were identified:

- a) *Residents in Gela*: workers born in Gela;

- b) *Moved to Gela* when hired: workers born in Sicilian municipalities with the probability of commuting defined by the model as <0.5

- c) *Commuters*: workers born in Sicilian municipalities with the probability of commuting defined by the model as ≥ 0.5

The study population included 5,627 workers born in Sicily and employed in the petrochemical complex from 1960 – the year plant operations started – to 1993. The vital status follow-up was from 1960 to 2002.

Mortality Rate Ratios (RRs) for all causes, all neoplasms and lung cancer, were estimated comparing workers by residence category. A Poisson regression model, controlling the RR for age and calendar period was applied using STATA 11.0 software; Confidence Intervals, CI (90%) were estimated by the maximum likelihood method. A regression model was also performed adding the variable of job title (results not reported); RR estimates did not significantly change (less than 8% for lung cancer) in the sensitivity analysis made by attributing different job categories to the subjects without information on job title.

Table 4 Classification decision rule with 0.5 probability cut-off

Sensitivity	n. of true “commuters”/(n. of true “commuters” + n. of false “no commuters”)	74.58%
Specificity	n. of true “no commuters”/(n. of “no commuters” + n. of false “commuters”)	92.99%
False Positive Rate	percentage of predicted commuters which are incorrect	17.76%
False Negative Rate	percentage of predicted no commuters which are incorrect	10.64%
Overall percentage correct		87.40%

Table 3 Distance thresholds

Labour force categories	Distance thresholds (km)
<i>lf₁</i> (Very Low labour force)	87.5
<i>lf₂</i> (Low labour force)	88
<i>lf₃</i> (Medium labour force)	97
<i>lf₄</i> (High labour force)	155.3

Descriptive data on workers by residence category are reported in Table 5.

Comparing mortality by place of birth and using *commuters* as a reference, the RR for lung cancer is 1.71 (0.92-3.17) for workers classified as *moved to Gela*, and 1.7 (1.03-2.81) for workers born in Gela; results are reported in Table 6.

The results obtained by applying the proposed categorization of residence are coherent with our hypothesis on residential risk for lung cancer mortality: a major risk is associated with residence in Gela.

RR results could be partly attributed to differences in distribution of other risk factors for lung cancer between residential categories, such as differences in smoking habits, which is the main recognized risk factor for lung cancer [86]. Though, examples of substantial confounding are rare in occupational epidemiology [64].

Despite some limitations, the results support the hypothesis that *commuters* have a lower residential/environmental risk of mortality from lung cancer compared with workers who were likely to have been residents in Gela.

Discussion and Conclusions

In this paper a multiple logistic regression model has been implemented for the estimation of distance thresholds to be applied in order to classify Gela's petrochemical workers in the following categories: "commuters", "moved to" and "residents". As in many retrospective studies, information from employment rosters about

Table 5 Descriptive data on Sicilian workers by residence category

	Commuters	Moved to Gela	Gela
N (workers)	3,234	709	1,684
<i>Causes of death</i>	N	N	N
All deaths	342	76	145
Neoplasms	101	23	53
Lung cancer	24	10	20
<i>Employment & follow-up</i>	Mean (SD)	Mean (SD)	Mean (SD)
Age at employment	26.1 (5.7)	26.3 (5.4)	26.6 (6.6)
Age at the end of follow-up	58.3 (9.6)	59.7 (9.9)	56.6 (10.2)
Latency ^a	32.2 (9.2)	33.4 (9.7)	30 (9.2)
<i>Job title</i>	N (%)	N (%)	N (%)
Blue collar	1,665 (51.5)	253 (37.7)	1,067 (63.4)
White collar	727 (22.5)	263 (37.1)	188 (11.2)
Both	608 (18.8)	142 (20)	322 (19.1)
Missing	234 (7.2)	51 (7.2)	107 (6.3)

^aPeriod from hiring to the end of follow-up (duration of follow-up).

Table 6 Rate Ratio (RR) of mortality from all causes, all neoplasms and lung cancer by residential categories using commuters as a reference

Cause of death (IX ICD ^a)	Category of residence	RR ^b	90% CI
All causes (001-999)	Commuters	1.0	-
	Moved to Gela	0.9	0.73-1.11
	Residents in Gela	0.88	0.75-1.03
All neoplasms (140-208)	Commuters	1.0	-
	Moved to Gela	0.94	0.64-1.37
	Residents in Gela	1.09	0.82-1.44
Lung cancer (162)	Commuters	1.0	-
	Moved to Gela	1.71	0.92-3.17
	Residents in Gela	1.7	1.03-2.81

^a International Classification of Disease IX revision codes.

^b RR adjusted for age and calendar period.

workers' place of residence are not reliable and reconstruction of residential history is not possible.

In similar situations, we suggest classifying population according to birthplace information, implementing the following procedure: a) using external data on workers' mobility (e.g., census data), construct a dichotomous variable for municipalities according to the presence of workers commuting to the destination under study (Gela, in our case), attributing value 1 in the case of positive mobility and value 0 in the case of no commuters; b) classify municipalities according to the dimension of their labour force, grouping count data into quartiles in order to consider the role of "economic mass" of localities; c) implement a logistic regression model to estimate a "probability of commuting" as a function both of distance and of four dummies accounting for the "economic mass" of the origin municipality; d) define "commuting origins" all municipalities where the probability of commuting is equal to or higher than 0.5 and, consequently, establish the "commuting threshold" at the distance that corresponds to a probability of 0.5 for each dimensional group of municipalities; e) classify workers that were born outside the commuting threshold as "moved to", workers that were born within the threshold as "commuters", and workers that were born in the destination municipality as "residents"; f) analyze data according to the resulting classification.

As far as the mobility analysis is concerned, the study is subject to implicit assumptions and restrictions:

- 1) The cohort study regards a specific group of workers (petrochemical workers), while mobility analysis is conducted on the labour force as a whole. This implies that petrochemical workers have the same commuting behaviour as the total Gela labour force.

2) The cohort study refers to the population that had been hired at the petrochemical complex during the 1960-1993 period, while the mobility analysis was based upon 2001 census observations. This implicitly assumes that the commuting habits of the Gela labour force (and of the petrochemical cohort) remained constant over the period 1960-2001.

3) Mobility analysis does not account for differences in accessibility conditions between the origin municipalities. Further developments should consider the distinction between different road characteristics (principal and secondary) and control for other modal conditions (railways).

The definition of residential categories obtained from the analysis of the observed mobility patterns, reduces misclassification of residential status for a qualitative categorisation, in the absence of the individual data necessary to correctly define a residential history for each subject. Further research should reveal more about the impact of commuting on exposure to environmental risk factors for populations living/working within contaminated areas.

Finally, as no ad hoc assumption is required, the methodology can be replicated to different contexts. Verification of mobility phenomena using qualitative sociological studies is suggested. Furthermore, a careful consideration of the specific characteristics of the regional context ("insularity" and its impact on transportation time and costs, in our case) is advised when specifying the area of application for the mobility analysis.

Acknowledgements

This study originates within a wider research project held by WHO-European Centre for health and environment (Rome) to assist the Sicilian regional administration in the management of risk areas; the research was financially supported by Italian Ministry of University and Research (PRIN 2008 Research Project: Metodi Economici e statistici per le ricerche di epidemiologia ambientale). The content of the paper reflects the authors' opinions and does not involve any institutional responsibility. The authors thank anonymous referees for useful comments and suggestions.

Author details

¹Department of Economics, Statistics, Mathematics and Sociology, University of Messina, Via Tommaso Cannizzaro 278, Messina, Italy. ²Department of Environment and Primary Prevention, Istituto Superiore di Sanità, Viale Regina Elena 299, Rome, Italy. ³World Health Organization - European Centre for Environment and Health, Via Francesco Crispi 10, 00187 Rome, Italy.

Authors' contributions

GS collaborated on the development of the idea for the study, on the definition of the mobility model supervising the research; RPo undertook the mortality analysis providing the workers' cohort data and the results section; EG worked on data acquisition and contributed to the mobility analysis providing the commuting probability sub-section; MM performed the statistical analysis for specifying commuting probabilities and prepared the mobility model results sub-section; MLR provided feedback on commuting study design and interpretation of results. PM cooperated to the

background analysis as far as the study area is concerned, providing figures and maps. All authors read and approved the final manuscript.

Conflict of interests

The authors declare that they have no competing interests.

Received: 7 October 2010 Accepted: 27 January 2011

Published: 27 January 2011

References

1. Sheaff R: Healthcare access and mobility between the UK and other European Union states: an 'implementation surplus'. *Health Policy* 1997, **42**:239-253.
2. Spicer N: Pastoral mobility, sedentarization and accessibility of health services in the northeast Badia of Jordan. *Applied Geography* 1999, **19**:299-312.
3. Mocellin J, Foggini P: Health status and geographic mobility among semi-nomadic pastoralists in Mongolia. *Health & Place* 2008, **14**:228-242.
4. Wallace R, Wallace D, Ullmann JE, Andrews H: Deindustrialization, inner-city decay and the hierarchical diffusion of AIDS in the USA: how neoliberal and cold war policies magnified the ecological niche for emerging infections and created a national security crisis. *Environment and Planning A* 1999, **31**:113-139.
5. Cliff A, (Eds): *Island Epidemics* Oxford University Press; 2000.
6. Dibble C: The GeoGraph 3D computational laboratory: Network and Terrain landscapes for RePast. *Journal of Artificial Societies and Social Simulation* 2004, **7**(Suppl 1).
7. Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N: Modelling disease outbreaks in realistic urban social networks. *Nature* 2004, **429**:180-184.
8. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A: Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(Suppl 51):21484-21489.
9. Boyle PJ, Gatrell AC, Duke-Williams O: The effect on morbidity of variability in deprivation and population stability in England and Wales: an investigation at small-area level. *Social Science & Medicine* 1999, **49**:791-799.
10. Boyle PJ, Gatrell AC, Duke-Williams O: Do area-level population change, deprivation and variations in deprivation affect individual-level self-reported limiting long-term illness? *Social Science & Medicine* 2001, **53**:795-799.
11. Davey Smith G, Shaw M, Dorling D: Shrinking areas and mortality. *The Lancet* 1998, **352**:1439-1440.
12. Bentham G: Migration and morbidity: Implications for geographical studies of disease. *Social Science & Medicine* 1998, **26**(Suppl 1):49-54.
13. Findley SE: The directionality and age selectivity of the health-migration relation: Evidence from sequences of disability and mobility in the United States. *International Migration Review* 1988, **22**:4-29.
14. Verheij RA, van de Mheen HD, de Bakker DH, Groenewegen PP, Mackenbach JP: Urban-rural variations in health in the Netherlands: Does selective migration play a part? *Journal of Epidemiology & Community Health* 1998, **52**:487-493.
15. Larson A, Bell M, Young AF: Clarifying the relationship between health and residential mobility. *Social Science & Medicine* 2004, **59**:2149-2160.
16. Brown D, Leyland AH: Population mobility, deprivation and self-reported limiting long-term illness in small areas across Scotland. *Health & Place* 2009, **15**:37-44.
17. Martikainen P, Sipilä P, Blomgren J, Van Lenthe FJ: The effects of migration on the relationship between area socioeconomic structure and mortality. *Health & Place* 2008, **14**:361-366.
18. Rogerson PA, Han D: The effects of migration on the detection of geographic differences in disease risk. *Social Science & Medicine* 2002, **55**:1817-1828.
19. Jacquez G, Meliker J, AvRuskin G, Goovaerts P, Kaufmann A, Wilson A, Nriagu J: Case-control geographic clustering for residential histories accounting for risk factors and covariates. *International Journal of Health Geographics* 2006, **5**:32.
20. Jacquez G, Meliker J, Kaufmann A: In search of induction and latency periods: Space-time interaction accounting for residential mobility, risk

- factors and covariates. *International Journal of Health Geographics* 2007, **6**:35.
21. Alm S, Mukala K, Jantunene MJ: **Personal carbon monoxide exposures of preschool children in Helsinki, Finland: levels and determinants.** *Atmospheric Environment* 2000, **30**:277-285.
 22. Punpuing S, Ross H: **The human side of Bangkok's transport problems.** *Cities* 2001, **18**:43-50.
 23. Batterman S, Peng C, Braun J: **Levels and composition of volatile organic compounds on commuting routes in Detroit, Michigan.** *Atmospheric Environment* 2002, **36**:6015-6030.
 24. Aarnio P, Yli-Tuomi T, Kousa A, Mäkelä T, Hirsikko A, Hämeri K, Räisänen M, Hillamo R, Koskentalo T, Jantunen M: **The concentrations and composition of and exposure to fine particles (PM_{2.5}) in the Helsinki subway system.** *Atmospheric Environment* 2005, **39**:5059-5066.
 25. Scotto di Marco G, Kephapopoulos S, Ruuskanen J, Jantunen M: **Personal carbon monoxide exposure in Kelsinki, Finland.** *Atmospheric Environment* 2005, **39**:2697-2707.
 26. Fondelli M, Chellini E, Yli-Tuomi T, Cenni I, Gasparri A, Nava S, Garcia-Orellana I, Lupi A, Grechi D, Mallone S, Jantunen M: **Fine particle concentrations in buses and taxis in Florence, Italy.** *Atmospheric Environment* 2008, **42**:8185-8193.
 27. Frumkin H: **Urban Sprawl and Public Health.** *Public Health Reports* 2002, **117**:201-217.
 28. Kageyama T, Nishikido N, Kobayashi T, Kurokawa Y, Kabuto M: **Commuting, overtime, and cardiac autonomic activity in Tokyo.** *The Lancet* 1997, **350**:693.
 29. Barengo N, Hu G, Lakka T, Pekkarinen H, Nissinen A, Tuomilehto J: **Low physical activity as a predictor for total and cardiovascular disease mortality in middle-aged men and women in Finland.** *European Heart Journal* 2004, **25**:2204-2211.
 30. Hamer K, Chida Y: **Active commuting and cardiovascular risk: A meta-analytic review.** *Preventive Medicine* 2008, **46**:9-13.
 31. Lerer LB, Matzopoulos R: **Meeting the challenge of railway injury in a South African city.** *The Lancet* 1996, **348**(Suppl 9028):664-666.
 32. Salminen S: **Traffic accidents during work and work commuting.** *International Journal of Industrial Ergonomics* 2000, **26**:75-85.
 33. Oja P, Vuori I, Paronen O: **Daily walking and cycling to work: their utility as health-enhancing physical activity.** *Patient Education and Counseling* 1998, **33**:587-594.
 34. Luoto R, Latikka P, Pukkala E, Hakulinen T, Vihko V: **The effect of physical activity on breast cancer risk: A cohort study of 30,548 women.** *European Journal of Epidemiology* 2001, **16**:973-980.
 35. Hu G, Jousilahti P, Borodulin K, Barengo N, Lakka T, Nissinen A, Tuomilehto J: **Occupational, commuting and leisure time physical activity in relation to coronary heart disease among middle-aged Finnish men and women.** *Atherosclerosis* 2007, **194**:490-497.
 36. Robertson-Wilson J, Leatherdale S, Wong S: **Social-Ecological Correlates of Active Commuting to School Among High School Students.** *Journal of Adolescent Health* 2008, **42**:486-495.
 37. Abu-Omar K, Rütten A: **Relation of leisure time, occupational, domestic, and commuting Physical activity to health indicators in Europe.** *Preventive Medicine* 2008, **47**:319-323.
 38. Merom D, Tudor-Locke C, Bauman A, Rissel C: **Active commuting to school among NSW primary school children: implications for public health.** *Health & Place* 2006, **12**:678-687.
 39. Cooper A, Wedderkopp N, Jago R, Kristinen P, Moller N, Froberg K, Page AS, Andersen LB: **Longitudinal associations of cycling to school with adolescent fitness.** *Preventive Medicine* 2008, **46**:324-328.
 40. Hume C, Timperio A, Salmon J, Craver A, Giles-Conti B, Crawford D: **Walking and Cycling to School. Predictors of Increases Among Children and Adolescents.** *American Journal of Preventive Medicine* 2009, **36**(Suppl 3):195-200.
 41. Setton EM, Keller CP, Cloutier-Fisher D, Hystad PW: **Spatial variation in estimated chronic exposure to traffic-related air pollution in working populations: a simulation.** *International Journal of Health Geographics* 2008, **18**:7-39.
 42. Scott L, Getis A: **Changing urban spatial structure and the jobs housing balance: The case of Los Angeles.** *Paper presented at the Western Regional Science Association annual meeting* Monterey, CA; 1998.
 43. Wang F: **Job proximity and accessibility for workers of various wage groups.** *Urban Geography* 2003, **24**(Suppl 3):253-271.
 44. Horner MW: **Spatial dimensions of urban commuting: A review of major issues and their implications for future geographic research.** *The Professional Geographer* 2004, **56**(2):160-173.
 45. Rouwendal J, Nijkamp P: **Living in two worlds: A review of home-to-work decisions.** *Growth and Change* 2004, **35**:287-303.
 46. Terracini B: **Aree oggetto di bonifica: inquadramento teorico e metodologico.** *ISTISAN Reports* 2005, **1**:53-61.
 47. Bosco ML, Varrica D, Dongorrà G: **Case study: inorganic pollutants associated with particulate matter from an area near a petrochemical plant.** *Environmental Research* 2005, **99**:18-30.
 48. Manno E, Varrica D, Dongorrà G: **Metal distribution in road dust samples collected in an urban area close to a petrochemical plant at Gela, Sicily.** *Atmospheric Environment* 2006, **40**:5929-41.
 49. Settimo G, Mudu P, Viviano G: **L'inquinamento atmosferico: problematiche generali nell'area di Gela.** *Epidemiologia & Prevenzione* 2009, **33**(Suppl 3):37-42.
 50. Musmeci L, Carere C, Falleni F: **Environmental pollution in the area of Gela.** *Epidemiologia & Prevenzione* 2009, **33**(Suppl 3):18-23.
 51. Bhopal RS, Moffatt S, Pless-Mulloli T, Phillimore PR, Foy C, Dunn CE, Tate JA: **Does living near a constellation of petrochemical, steel, and other industries impair health? Occupational and Environmental Medicine** 1998, **55**:812-822.
 52. Edwards R, Pless-Mulloli T, Howel D, Chadwick T, Bhopal R, Harrison R, Gribbin H: **Does living near heavy industry cause lung cancer in women? A case-control study using life grid interviews.** *Thorax* 2006, **61**:1076-82.
 53. Gottlieb MS, Shear CL, Seale DB: **Lung cancer mortality and residential proximity to industry.** *Environmental Health Perspective* 1982, **45**:157-164.
 54. Yang CY, Cheng MF, Chiu JF, Tsai SS: **Female lung cancer and petrochemical air pollution in Taiwan.** *Archives of Environmental Health* 1999, **54**:180-185.
 55. Belli S, Benedetti M, Comba P, Lagravinese D, Martucci V, Morleo D, Trinca S, Viviano G: **Case-control study on cancer risk associated to residence in the neighbourhood of a petrochemical plant.** *European Journal of Epidemiology* 2004, **19**:49-54.
 56. Checkoway H, Pearce N, Kriebel D: *Research Methods in Occupational Epidemiology* New York: Oxford University press; 2004.
 57. Martuzzi M, Mitis F, Biggeri A, Terracini B, Bertollini R: **Environment and health status of the population in areas with high risk of environmental crisis in Italy.** *Epidemiologia & Prevenzione* 2002, **26**(Suppl 6):1-53.
 58. Fano V, Cernigliaro A, Scondotto S, Addario SP, Caruso S, Mira A, Forastiere F, Perucci CA: **Mortality (1995-2000) and hospital admissions (2001-2003) in the industrial area of Gela.** *Epidemiologia & Prevenzione* 2006, **30**(Suppl1):27-32.
 59. Cernigliaro A, Pollina Addario S, Cesaroni G, Fano V, Fantaci G, Tavormina E, Marras A, Dardanoni G, Forastiere F, Perucci CA, Scondotto S: *Stato di salute nelle aree a rischio ambientale della Sicilia. Aggiornamento dell'analisi della mortalità (anni 1995-2002) e dei ricoveri ospedalieri (anni 2001-2006)* Palermo: Dipartimento Osservatorio Epidemiologico, Assessorato alla Sanità, Regione Sicilia; 2008 [http://www.doescilia.it/media/RapportoAreeARischio_al2006.pdf].
 60. La Rocca M: *Modelli di industrializzazione e salute umana. Lo sviluppo tra benessere e rischio nei poli petrolchimici della Sicilia* Milano: Franco Angeli Editore; 2010.
 61. Saitta P: *Spazi e società a rischio. Ecologia, petrolio e mutamento a Gela* Napoli: Think Thanks; 2009.
 62. Pasetto R, Comba P, Pirastu R: **Lung cancer mortality in a cohort of workers in a petrochemical plant: occupational or residential risk?** *International Journal of Occupational and Environmental Health* 2008, **14**:124-128.
 63. Pearce N, Checkoway H, Kriebel D: **Bias in occupational studies.** *Occupational and Environmental Medicine* 2007, **64**:562-568.
 64. Blair A, Stewart P, Lubin JH, Forastiere F: **Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures.** *American Journal of Industrial Medicine* 2007, **50**:199-207.
 65. Sen A, Sööt S: **Selected procedures for calibrating the generalized gravity model.** *Papers of the Regional Science Association* 1981, **48**:165-176.
 66. Sen A, Smith TE: *Gravity Models of Spatial Interaction Behavior* Berlin: Springer; 1995.

67. De Vries JJ, Nijkamp P, Rietveld P: **Exponential or power distance-decay for commuting? An alternative specification.** Tinbergen Institute; 2004 [<http://www.tinbergen.nl>], Discussion Paper 097/3.
68. Fotheringham AS, O'Kelly ME: *Spatial Interaction Models: Formulations and Applications* Netherlands: Kluwer Academic Publishers; 1989.
69. Smith TS: **Inverse distance variations for the flow of crime in urban areas.** *Social Forces* 1976, **54**(Suppl 4):802-815.
70. Pellegrini PA, Fotheringham S: **Modelling spatial choice: a review and synthesis in a migration context.** *Progress in Human Geography* 2002, **26**:487-510.
71. Haynes RM: **Application of exponential distance decay to human and animal activities.** *Geografiska Annaler: Series B, Human Geography* 1974, **56**(Suppl 2):90-104.
72. Nekola JC, White PS: **The distance decay of similarity in biogeography and ecology.** *Journal of Biogeography* 1999, **26**:867-878.
73. Hanley N, Schläpfer F, Spurgeon J: **Aggregating the benefits of environmental improvements: distance-decay functions for use and non-use values.** *Journal of Environmental Management* 2003, **68**:297-304.
74. Feenstra RC, Markusen JA, Rose AK: **Using the gravity equation to differentiate among alternative theories of trade.** *Canadian Journal of Economics* 2001, **34**(Suppl 2):430-447.
75. Krugman P, Obstfeld M: *International Economics: Theory and Policy* U.S.: Pearson Education; 2008.
76. Ogura LM: **Effects of urban growth controls on intercity commuting.** *Urban Studies* 2010, **47**(10):2173-93.
77. De Vries JJ, Nijkamp P, Rietveld P: **Estimation of Alonso's theory of movements by means of instrumental variables.** *Network and Spatial Economics* 2002, **2**(Suppl 2):107-126.
78. Morgenroth ELW: **Commuting in Ireland: An analysis of inter-county commuting flows.** Economic and Social Research Institute; 2002, WP144 [<http://www.esri.ie/pdf/COMMUTING.PDF>].
79. Guagliardo MF: **Spatial accessibility of primary care: concepts, methods and challenges.** *International Journal of Health Geographics* 2004, **26**(3):3.
80. Bharti N, Xia Y, Bjornstad ON, Grenfell BT: **Measles on the edge: coastal heterogeneities and infection dynamics.** *PLoS-One* 2008, **3**(4):e1941.
81. Haynes KE, Fotheringham AS: *Gravity and Spatial Interaction Models* Beverly Hills: Sage Publications; 1984.
82. Flowerdew R, Aitkin M: **A method of fitting the gravity model based on the Poisson distribution.** *Journal of Regional Science* 1982, **22**:191-202.
83. Flowerdew R, Amrhein C: **Poisson regression models of Canadian census division migration flows.** *Papers in Regional Science* 1989, **67**(Suppl 1):89-102.
84. Sarra AL, Del Signore M: **A dynamic origin-constrained spatial interaction model applied to Poland's inter-provincial migration.** *Spatial Economic Analysis* 2010, **5**(Suppl 1):29-41.
85. Istat: *Censimento Generale della Popolazione* Roma; 2001.
86. Alberg AJ, Ford JG, Samet JM, American College of Chest Physicians: **Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition).** *Chest* 2007, **132**:295-555.

doi:10.1186/1476-072X-10-11

Cite this article as: Signorino et al.: Gravity models to classify commuting vs. resident workers. An application to the analysis of residential risk in a contaminated area. *International Journal of Health Geographics* 2011 **10**:11.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

