An Atlas of Cancer in Ireland 1994-2003

Anne-Elie Carsin

Linda Sharp

Harry Comber

National Cancer Registry
Building 6800
Cork Airport Business Park
Cork

Tel +353 21 4318014
fax +353 21 4318016
Email: info@ncri.ie
Web www.ncri.ie

Published by the National Cancer Registry
ISBN 978-0-9554970-1-8

Summary

Background

Many of the causes of cancer are still not well understood. Although much is known about the effects of lifestyle and environmental factors, a considerable part of the cancer burden remains unexplained. Investigation of geographical differences in cancer incidence can throw light on both cancer aetiology and also on variations in cancer risk factors between populations. Spatial variation on a relatively fine scale may also yield information on the success of programmes of prevention, screening and early detection.

The aim of this report was to describe variations in cancer risk (incidence) at electoral district (ED) level in Ireland. The objectives were to:

- investigate geographical variation in cancer incidence in Ireland;
- examine the relationships between area-based characteristics (such as population density) and cancer risk;
- attempt to explain these relationships through the examination of area-based measures of socio-economic status and aspects of lifestyle.

Methods

The analyses were based on cancers diagnosed in the population of Ireland during 1994-2003, and registered with the National Cancer Registry. Each case was assigned to an electoral division (ED), based on the address of the patient at the time of diagnosis. The ED was used to allocate a value to each case, for a range of area-based measures of socio-economic status. Cases were assigned to a deprivation category, ranging from least (level 1) to most (level 5) deprived, based on the deprivation index developed by the Small Area Health Research Unit from various 2002 census socio-economic variables. A measure of the population density of each ED was created, based on the average number of inhabitants at the 1996 and 2002 censuses. EDs were combined into approximate tertiles for analysis (<1 person/hectare, 1-20 persons/hectare, >20 persons/hectare) and cases assigned to the appropriate tertile. EDs were also aggregated into quartiles of a range of socio-economic variables from the 2002 census: % unemployed, % agricultural workers, % lower social class, % manual workers, % non-manual workers, % early school leavers, % with no car, % local authority housing, % overcrowded housing and % of persons aged 65 and older living alone. Cases were assigned to the appropriate quartile for each variable. Population data was derived from the census Small Area Population Statistics (SAPS) files for 1996 and 2002.

In the spatial analysis, for each cancer site, an age-standardised incidence ratio (SIR) was computed for each ED. Bayesian conditional autoregressive models (CAR) were used to smooth these estimates. Models were fitted using the Gibbs Markov Chain Monte Carlo algorithm in WinBUGS. The smoothed risk estimates (relative risks, RRs) were mapped for each cancer site individually. For those cancers which affect both sexes, relative risks were mapped for both sexes combined. and for males and females separately.

Poisson regression was used to investigate the relationships between the risk of cancer and deprivation, population density and the other area-based socio-economic variables. In each analysis, the lowest quantile was taken as the reference group. Relative risks for deprivation were adjusted for population density; risks for density and other socio-economic variables which were significantly associated with cancer incidence were mutually adjusted.

Data from the SLÁN survey on various aspects of socio-economic status, diet and lifestyle (e.g. % low income, % current smokers, etc) was mapped at the level of rural districts and informally compared to the cancer incidence maps where relevant.

Results

Geographical variation

- **All malignant cancers (excluding non-melanoma skin cancer):** In both men and women, there were areas of higher incidence around Dublin and Cork and, for men, around some other urban centres. Incidence was also higher than average in a band running across the northeast and north midlands, from Dublin to Sligo.
- **Non-melanoma skin cancer:** The geographical distribution of non-melanoma skin cancer was similar in men and women but the variation was somewhat more pronounced for men. Areas of higher incidence were seen around the cities of Dublin, Cork, Galway and Waterford. Within Cork and Dublin, the areas of higher incidence were in the south and east of the cities, respectively. Outside the urban areas, regions of high incidence were observed in areas along the west coast of Donegal, Mayo, Clare, Kerry, west Cork (men) and also on the coast of Waterford (men).
- **Breast cancer:** There was relatively modest geographical variation in breast cancer incidence. The areas of highest incidence were around the major urban areas, with the exception of Limerick. There was a slightly increased incidence in west Cork, north Kerry, and a large area in the east Midlands. Within Dublin, incidence was higher in the southeast than in the north and west.
- **Colorectal cancer:** There was evidence of moderate geographical variation in colorectal cancer incidence. Incidence was higher than average in two areas - one centred on Cork city but extending into the far southwest - and the other in the north and centre of the country, in a broad band from Dublin through the northeast to Donegal. The pattern was similar in both sexes although for women incidence was higher in the centre and the northwest.
- **Lung cancer:** In both sexes, there was an area of higher lung cancer incidence in Leinster, with the highest rates in Dublin, Kildare and Wicklow. A much smaller area of high incidence was centred on Cork city. For men, there were pockets of high incidence in the northwest, in Sligo, Leitrim and Donegal. Within Dublin and Cork, the areas of highest incidence coincided with the more deprived areas in the north and northwest, respectively.

- **Prostate cancer:** Prostate cancer incidence was highest around the major urban centres, with the exception of Limerick. Within Dublin, incidence was higher in the south of the city than in the north. There were also distinct areas of higher incidence in the northwest of the country, in Sligo and Donegal.
- **Stomach cancer:** Stomach cancer showed one of the strongest patterns of geographical clustering, with higher incidence in two clearly defined areas; one covering the northeast, stretching from Dublin through Louth, Monaghan and Cavan, and the other in south Donegal. Within Dublin, incidence was highest in the north and west of the city. The pattern was quite similar in both sexes.
- **Bladder cancer:** Geographical variation in bladder cancer was more marked in men than women. In men, there were three areas of higher incidence - along the east coast in Dublin and Wicklow, in Co. Donegal, and around Cork city. The pattern for women was less distinct, but there were again areas of higher incidence around Dublin (mainly confined to the city) and in Donegal, confined mainly to the Inishowen peninsula, and a trend of slightly increasing incidence heading towards the southwest.
- **Melanoma of the skin:** There were pronounced areas of higher incidence in west Cork, in, and to the north of, Dublin, in and around Cork and Waterford, and along the west coast of Donegal. Among men, there were also some patches of higher incidence in the west, on the coasts of Co. Galway and Co. Mayo. Within Dublin, incidence was highest in the south of the city.
- **Head and neck cancer:** For men, there were several patches of high incidence - in the main urban centres, in a band running from Cork to Galway, in a broad area in the north midlands, in northwest Mayo and in the Iveragh peninsula in Kerry. Within Cork and Dublin, head and neck cancer was more common in more deprived areas. In women, geographical variation was less marked. There was a region of higher incidence in and around Dublin and in the northeast, with a smaller area with higher rates in the northeast tip of Co. Donegal.
- **Oesophageal cancer:** Few areas had a particularly high incidence of oesophageal cancer. The country was split into areas of lower incidence in the northwest of the country (Galway, Clare, Sligo and Donegal counties) and those of slightly higher incidence in the northeast and running toward the south and west.
- **Cancer of the cervix uteri:** The areas of highest incidence of cervical cancer were concentrated in and around Dublin and in a broad band down the eastern side of the country from Dublin through Kildare and Wicklow to Wexford. There was another less concentrated band of higher incidence running through the middle of the country from north to south. Lower incidence was observed in the southwest, in counties Cork and Kerry, as well as in Donegal in the northwest.

Deprivation

All of the cancer sites analysed showed some association with deprivation, either an increase with increasing deprivation (all malignant cancers and colorectal, lung, stomach, bladder, head and neck, cervical and oesophageal cancer) or a decrease (breast, prostate and non-melanoma and melanoma skin cancers). In general, the relative risk estimates for the most, compared to the least, deprived were relatively modest, falling in the range 0.8-1.3. Stronger associations were seen for lung cancer in men (RR=1.72) and women (RR=1.56), head and neck cancer in men (RR=1.78), cervical cancer (RR=1.74), and melanoma (RR in both sexes 0.64-0.66).

Population density

With the exception of prostate cancer, all of the cancers considered in this report were significantly associated with population density. More densely populated areas (those with a population of >20 persons/hectare) consistently had a higher risk of cancer than those that were sparsely populated (<1 persons/hectare). Some of the observed associations were reasonably strong: relative risks were 1.4 or higher for cancers of the bladder (men, RR=1.39; women RR=1.40), stomach (men, RR=1.45; women, RR=1.49) and lung (men, RR=1.62; women, RR=1.84).

Other area-based measures of socio-economic status

With the exception of cervical cancer, the risk of all cancers analysed in this report was higher in areas with the highest proportion of elderly people living on their own. Although the risk estimates were less than 1.3, this association between this factor and almost every cancer was statistically significant.

Areas with a higher percentage of agricultural workers had a consistently lower risk of cancer. This was seen for all cancers with the exception of prostate cancer.

The observed relationships between the other area-based characteristics and cancer risk - such as percentages of lower social class, unemployed, living in overcrowded housing, and early school leavers - tended to mirror the associations with deprivation.

Discussion

There are geographical variations in the risk of cancer across Ireland. For some cancers these patterns are quite striking (e.g. lung cancer, cervical cancer, non-melanoma skin cancer, melanoma of the skin), while for others they are less marked (e.g. breast cancer). Although some similarities were apparent (e.g. between lung cancer and other smoking-related cancers, between non-melanoma cancer and melanoma of the skin, and between breast and prostate cancer), the observed geographical variations were, in the main, different for different cancers. Generally, for those cancers that affect both sexes, the geographical distribution was similar for men and women.

It must be kept in mind that these variations in risk do not mean that the spatial location itself causes cancer; rather they are likely to reflect socio-economic differences in the population, geographical differences in exposure to risk factors and, for some cancer sites, variations in access to, or uptake, of screening or other cancer services.

As regards deprivation, the observed associations between deprivation and cancer incidence in Ireland are generally consistent with those reported from other countries, using both area-based measures of deprivation and a range of other individual-level measures of socio-economic status (e.g. occupation, education, housing tenure, income). Socio-economic variations in several lifestyle risk factors for cancer (e.g. smoking) are well known, and these probably underlie the observed associations.

The associations between cancer incidence and population density are likely to be, in part, due to residual confounding by socio-economic status, at least for those cancers positively associated with deprivation. But this cannot be the entire explanation, and it is likely that there are urban/rural variations in exposure to cancer risk factors and in health behaviours, including health service access and utilisation.

The inverse associations between the percentage of agricultural workers and risk of several cancers are, most probably, a reflection of the relationship between cancer risk and population density.

The similar associations between cancer risk and (a) overall deprivation and (b) individual measures of socio-economic status, such as unemployment, was unsurprising since several of these individual factors are included in the composite deprivation index.

The consistent association with the proportion of elderly living alone is hard to interpret. It seems most likely that it either reflects differences in patterns of exposure to cancer risk factors in older people who live alone compared to those who live with others, or is a proxy for some other unmeasured cancer risk factor.

Conclusions

This report has revealed geographical and socio-economic variations in cancer risk in Ireland. These are likely to reflect differences in social, economic, cultural and environmental differences between subgroups of the population. Although risk factors for cancer are not all well-defined, nor modifiable (e.g. family history, genetic background), it is likely that many of the differences observed reflect a combination of variations in well-known risk factors (such as tobacco smoking, alcohol drinking, obesity, diet, sexual behaviour, etc.) and variations in participation in screening, health awareness and access to cancer services. Since these factors are potentially modifiable, there is considerable potential for reducing cancer incidence in Ireland and eliminating the disparities described in this report.

Contents

Index of maps

Acknowledgements

This report would not have been possible without the dedicated efforts of staff at the National Cancer Registry, and the support and help that they receive from hospital staff throughout the country. We are particularly grateful to the Tumour Registration Officers who collected the registrations on which this report was based, and to Neil McCluskey and Fiona Dwane who undertook the geocoding of the cancer cases which made the geographical analysis possible.

We would also like to thank several other organisations and individuals for providing data included in the report. The Central Statistics Office provided the small area census population data and the information from which numbers of deaths from cancer was derived. The SLÁN research group (http://www.slan06.ie/team.htm) very kindly agreed to share their data and we are indebted to them for this, in particular Dr Dorothy Watson, Senior Research Officer at the Economic and Social Research Institute. The map of radon exposure was provided by the Radiological Protection Institute of Ireland. Dr Alan Kelly and colleagues, from the Small Area Health Research Unit, Trinity College Dublin, created the deprivation index which was used in this report.

1 Introduction

1.1 Background

Cancer is now the most common cause of death in Ireland (Central Statistics Office Ireland, 2009). Overall cancer incidence is expected to increase by 45% between 2010 and 2020, and by 110% between 2010 and 2030 (National Cancer Registry, 2008), mainly due to population ageing. Cancer mortality is also projected to increase (National Cancer Registry, 2003) although not to the same extent.

Many of the causes of cancer are still not well understood. Although much is known about the effects of lifestyle and environmental factors (see, for example, World Cancer Research Fund / American Institute for Cancer Research 2007, Boyle and Levin, 2008), a considerable part of the cancer burden remains unexplained. Investigation of geographical differences in cancer incidence can throw light on both cancer aetiology and also on variations in cancer risk factors between populations. Spatial variation on a relatively fine scale may also yield information on the success of programmes of prevention, screening and early detection. In many countries, including Ireland, where information on personal characteristics of cancer patients is not available to cancer registries for legal reasons, information at small area level can act as a proxy for individual-level data, and can give valuable information on the role of diet, lifestyle, and socio-economic factors, on the cancer burden. Such data can also highlight disparities or variations in access to cancer services at all levels.

1.2 Aim of the report

Worldwide variation in cancer incidence has been extensively studied, most comprehensively in the quinquennial reports “Cancer Incidence in Five Continents” (Curado et al, 2007), produced by the International Agency for Research on Cancer, and publications based on this data (see, for instance, Bray et al, 2004, Bray et al, 2005, Devesa et al, 2005). Part of this variation is due to genetics, and the past few years have seen major advances in the understanding of the genetic and molecular basis of the disease. However, the majority of the variation is a result of social, economic, cultural and environmental differences between populations and describing variations in cancer rates between countries has served to provide clues to specific aetiological factors involved.

Variations in cancer risk and aetiological factors between countries are often large and readily amenable to study, but the study of the much smaller range of geographical variation within countries is more challenging. However, it also has the potential to provide insights which are of local significance. Although current cancer patterns reflect past patterns of exposure to risk factors, taking steps now to deal with these factors in the population has the potential to bring about reductions in future cancer incidence and mortality. Sometimes, merely drawing attention to variation can influence behaviour at both official and individual level to reduce cancer risk. Geographical variation in cancer incidence and mortality and survival (Kogevinas et al, 1997, Coleman et al, 1999) has been closely linked to patterns of socio-economic status and deprivation. Identification of these patterns can draw attention to the wider dimensions of health which need to be addressed in order to reduce cancer morbidity and mortality.

The aim of this report was to describe variations in cancer risk (incidence) at electoral district level in Ireland, with a view to identifying remediable risk factors. The objectives of the report were to:

- investigate geographical variation in cancer incidence in Ireland;
- examine the relationships between geographically-based characteristics (such as population density) and cancer risk;
- attempt to explain these relationships through the examination of area-based measures of socio-economic status and aspects of lifestyle.

1.3 Content of the report

This report brings together - for the first time - detailed descriptions of geographical variations in cancer risk in Ireland, with census data on characteristics of local areas and survey data on lifestyle factors. Cancer incidence rates across the country have been mapped using sophisticated methods of spatial analysis. The available data on risk factors has also been mapped, and statistical analysis has been used to explore links between the area characteristics and cancer risk.

Chapter 2 describes the data included in the report, and the methods of analysis. Chapter 3 includes results of the analysis for all malignant cancers, and chapters 4 to 14 include results for 11 of the most common cancer sites - namely non-melanoma skin, colorectal, breast, prostate, lung, stomach, bladder, head and neck and oesophageal cancer, cancer of the cervix uteri, and malignant melanoma of the skin. Chapter 15 contains incidence maps for six additional cancers (lymphoma, leukaemia and cancers of the pancreas, ovary, corpus uteri and brain and central nervous system (CNS)); these are presented in summary form because either the annual incidence was considered to be too low to justify full analysis, or there was little in the way of a geographical pattern. Maps showing the geographical distribution of selected cancer risk factors are included in Appendix 1. Appendix 2 contains summary tables from analyses of area characteristics (e.g. deprivation, population density) and cancer risk. Appendix 3 includes summary statistics related to the maps of cancer incidence. A map showing county boundaries in Ireland is provided in Appendix 4.

2 Methods

2.1 Data sources

2.1.1 Cancer registrations

The analyses in this report are based on cancers diagnosed in the population of Ireland during 1994-2003, and registered with the National Cancer Registry. Since 1st January 1994, all newly diagnosed cancers in Ireland have been registered by the National Cancer Registry. The process is highly effective, with over 96% of cancers being identified (National Cancer Registry, 2001). Prior to 1994, there was no national cancer registration and therefore no reliable information available on cancer incidence.

A summary of the cancers included in this report is given in table 2.1. Those tumours defined as "multiple primary cancers" according to international guidelines (Ferlay et al, 2005) were identified, and only a single instance of each cancer has been counted. When several primary malignant tumours occurred in the same site, only the first occurrence was considered. Cancer registration is a dynamic process and registrations may be added, changed or removed from the database over time as new information comes to light, sometimes several years after the original diagnosis. This means that the numbers of cancers in this report may differ slightly from those published elsewhere.

Table 2.1 Incident cancers diagnosed 1994-2003 and included in this report

Cancer site	ICD 10 codes	Total no. of cases, 1994-2003		Annual average no. of cases, 1994-2003	
		females	*males*	*females*	*males*
all malignant cancers	C00-C96	87,299	94,657	8,730	9,466
all malignant cancers, excl C44[1]	C00-C96, excl C44	64,002	68,519	6,400	6,852
non-melanoma skin	C44	23,297	26,138	2,330	2,615
breast	C50	18,196	128[2]	1,820	13[2]
colorectal	C18-C21	7,873	10,321	787	1,032
lung	C34	5,846	10,246	585	1,025
prostate	C61	-	15,252	-	1,525
lymphoma	C81-C85	2,433	2,853	243	285
stomach	C16	1,830	2,920	183	292
bladder	C67	1,320	3,312	132	331
melanoma of the skin	C43	2,659	1,624	266	162
leukaemia	C91-C95	1,602	2,292	160	229
head and neck	C01-C14, C30-C32	1,010	2,759	101	276
pancreas	C25	1,800	1,787	180	179
ovary	C56	3,454	-	345	-
brain and other central nervous system	C70-C72	1,450	1,835	145	183
kidney	C64	1,101	1,966	110	197
oesophagus	C15	1,205	1,861	120	186
corpus uteri	C54	2,332	-	233	-
cervix uteri	C53	1,834	-	183	-

[1] excludes non-melanoma skin cancer; [2] since breast cancer in males is rare, the analyses in chapter 5 are limited to breast cancer in females

2.1.2 Geocoding cancer cases to electoral divisions

The address of each cancer patient at the time of diagnosis is recorded by the Registry. The county of residence can be easily determined in the majority of cases from the address given. However, for detailed geographical analysis, each case must be assigned to an area much smaller than a county. In this way, areas of high cancer incidence can be more precisely defined and information on cancer incidence can be linked to known characteristics of an area, such as population density and deprivation (see below). The smallest useful area for this purpose in Ireland is the electoral division (ED) - formerly known as a district electoral division - as this is the smallest area for which census data can be obtained. These areas have a mean population of around 1,000 people, but can be much larger, and will typically be quite heterogeneous in population compared to, for instance, census enumeration districts in the UK (Coleman et al, 2001).

In theory, each cancer patient can be assigned to an ED, using the address given to the hospital at the time of diagnosis; this process is known as geocoding. However, in Ireland, addresses are not unique and have no postcodes, so they must be assigned to an ED by matching the address given to those in a database of all known addresses and their associated EDs. Three databases of this kind are available in Ireland - GeoDirectory, from An Post/OSI; address tables from the quinquennial censuses from the Central Statistics Office (CSO); and the electoral registers. All of these databases have limitations. None can be completely up-to-date, although the GeoDirectory is updated four times a year. GeoDirectory, in general, holds only one address, the official postal address, for each building, so alternative addresses, which are quite common in rural Ireland, are often not listed. The census tables are quite incomplete, and many addresses are not registered. The electoral registers were the most comprehensive listing of addresses, and tended to use addresses in everyday use rather than the postal address. However, with the passing of the Electoral (Amendment) Act, 2001, access to the full register was ended and the edited register now available is of much less value for geocoding. Using a combination of these three databases, it should be theoretically possible to match the addresses of all cancer patients to EDs. In practice, however, many addresses available to the Registry are incomplete, non-standard or inaccurate. In addition, none of the available databases lists every address and some have errors. At best, only 70% to 80% of addresses have a close match in any of the databases, and the remaining 20-30% have to be matched manually by inspection of individual records, with reference to large-scale maps. As the Registry records over 20,000 new cases each year, assigning an ED to each case is a time-consuming process.

As part of an ongoing geocoding project, the cancer cases included in this analysis were assigned to EDs using probabilistic matching software developed by the Registry specifically for this purpose. Addresses were also matched independently to the GeoDirectory database and to the electoral register for the same period. Addresses which could not be assigned to one specific ED by this process were individually inspected by Registry staff and, by referring to the GeoDirectory and the electoral registers, all but a small number could be allocated to an ED. For those registrations where a single ED could not be definitely assigned (3.9% of all malignant cancers; table 2.2), a number of alternative EDs were assigned. In calculating incidence rates for each ED (see below), a fraction of the cases was allocated to each of the alternative EDs. At the end of the process, a number of registrations remained which could not be assigned to any ED (4.6% of all malignant cancers; table 2.3). These registrations were excluded from the analyses in this report. This loss was taken into account in the calculation of the incidence rates (see below 2.1.3.1).

Apart from its obvious use in allocating cancer cases to specific areas and studying geographical patterns, geocoding provides a "key", allowing cancer cases to be linked to other area-based data, such as measures of socio-economic status (e.g. deprivation indices (Small Area Health Research Unit (SAHRU), 1997), percentage unemployed, etc) or population density. This is described in more detail below. This type of information is not, in general, accessible at the level of the individual cancer case in Ireland, and has to be inferred from area-based measures.

Table 2.2 Outcome of process of assigning cancer cases to EDs: cases not assigned to an ED and cases assigned to multiple EDs

Cancer site	Cases not assigned to an ED		Cases assigned to more than one ED		
	No. of cases	*% of cases*	*No. of cases*	*% of cases*	*No. of EDs*
all malignant cancers	8,422	4.6%	7,112	3.9%	2,279
all malignant cancers excl C44[1]	5,839	4.4%	5,091	3.8%	2,005
non-melanoma skin	2,580	5.2%	2,022	4.1%	1,608
breast	699	3.8%	640	3.5%	795
colorectal	755	4.1%	759	4.2%	912
lung	700	4.4%	514	3.2%	717
prostate	725	4.8%	622	4.1%	811
lymphoma	266	5.0%	187	3.5%	321
stomach	226	4.8%	201	4.2%	348
bladder	169	3.6%	179	3.9%	303
melanoma of the skin	231	5.4%	143	3.3%	270
leukaemia	181	4.6%	164	4.2%	296
head and neck	142	3.8%	139	3.7%	246
pancreas	165	4.6%	154	4.2%	285
ovary	157	4.5%	130	3.8%	225
brain and other CNS	161	4.9%	122	3.7%	231
kidney	143	4.7%	126	4.1%	227
oesophagus	141	4.6%	140	4.6%	243
corpus uteri	85	3.6%	69	2.9%	135
cervix uteri	74	4.0%	58	3.2%	100

[1] excludes non-melanoma skin cancer

2.1.3 Characteristics of EDs: population and socio-economic variables

2.1.3.1 Population

The 2002 census provided population data, broken down by age and sex, for 3,422 EDs in Ireland. These had an average population of 1,145; ranging from 55 (Branchfield, Co. Sligo) to 24,404 (Blanchardstown-Blakestown,

Table 2.3 Confidential electoral divisions - 2002

County	Confidential ED			ED combined with			New ED	
	No.	*Name*	*No. of persons (2002)*	*No.*	*Name*	*No. of persons (2002)*	*Name*	*No .of persons (2002)*
Laoighis	046	Capard	47	045	Brisha	224	Brisha/Capard	271
Longford	035	Newgrove	37	024	Firry	172	Firry/Newgrove	209
Offaly	034	Ballaghassaan	34	043	Esker	350	Esker/Ballaghassaan	384
Clare	017	Ballyeighter	46	020	Glenroe	117	Glenroe/Ballyeighter	163
Clare	133	Inishcaltra South	41	132	Inishcaltra North	276	Inishcaltra North/ Inishcaltra South	317
Cork	046	Whiddy	29	033	Bantry Rural	952	Bantry Rural/Whiddy	981
Tipperary North	045	Lackagh	20	037	Greenhall	237	Greenhall/Lackagh	257
Waterford City	006	Ballynaneashagh	17	002	Ballybeg South	265	Ballybeg South/ Ballynaneashagh	282
Waterford	074	Kilbarry (part)	45	070	Ballynakill (part)	327	Ballynakill (part)/Kilbarry (part)	372
Galway	027	Derrycunlagh	47	022	Bencorr	201	Bencorr/Derrycunlagh	248
Galway	126	Loughatorick	34	129	Marblehill	363	Marblehill/Loughatorick	397
Leitrim	034	Arigna	15	041	Garvagh	119	Garvagh/Arigna	134
Leitrim	029	Aghavoghill	40	027	Aghalateeve	94	Aghalateeve/Aghavoghill	134
Mayo	065	Sheskin	25	058	Glenco	100	Glenco/Sheskin	125
Mayo	130	Bundorragha	97	150	Owennadornaun	96	Owennadornaun/Bundorragha	193
Sligo	027	Mullagheruse	49	031	Templeboy South	194	Templeboy South/Mullagheruse	243
Cavan	082	Derrynananta	39	084	Dunmakeever	130	Dunmakeever/Derrynananta	169
Cavan	087	Teebane	34	086	Killinagh	110	Killinagh/Teebane	144
Cavan	028	Tircahan	32	025	Pedara Vohers	154	Pedara Vohers/Tircahan	186

Co. Dublin). The population of a number of EDs was so low that the CSO considered these EDs "confidential", only published total population figures for them, and amalgamated them with one or more neighbouring EDs. EDs were considered confidential if they included either 15 households or less or 50 persons or less. There were 19 such confidential EDs in 2002 and these are shown in table 2.3.

Population data was derived from the census Small Area Population Statistics (SAPS) files for 1996 and 2002. SAPS populations from the 1996 census were used as the denominators for cases incident in 1994-1996. Data from the 2002 census was used for cases incident in 2002 and 2003, and a linear interpolation of the 1996 and 2002 census counts was used for cases incident in 1997-2001.

The definition of a small number of EDs, and therefore the associated SAPS data, changed between the 1996 and 2002 censuses. These changes consisted of splitting or amalgamation of areas, rather than any movement of boundaries. EDs which had changed in this way were combined for analysis, and the available age and sex distribution similarly combined (table 2.4). This combining of areas gave a final total of 3,419 EDs.

Table 2.4 Combined EDs with boundary changes between 1996 and 2002 censuses

ED number	Geographical area	ED name	SAPS data 1996[1]	Published total figure 1996[2]	SAPS data 2002[1]	Published total figure 2002[2]
19003	Co 18 Cork County	Tralee U.D.	19,056	6,085	6,311	6,311
19165	Co 19 Kerry	Tralee Rural (part)	860	12,971	15,433	14,064
		Tralee Rural (part)		860		1,369
33003	Co 32 Cavan	Letterkenny U.D.	7,606	2,473	2,478	2,478
33105	Co 33 Donegal	Letterkenny Rural (part)	2,341	5,133	9,289	5,487
		Letterkenny Rural (part)		2,341		3,802
34004	Co 33 Donegal	Monaghan U.D.	5,628	2,014	2,032	2,032
34063	Co 34 Monaghan	Monaghan Rural (part)	1,207	3,614	4,969	3,685
		Monaghan Rural (part)		1,207		1,284

[1] source: SAPS files where population data is available by age group and sex; [2] source: Central Statistics Office, 2003

2.1.3.2 Population density

As the formal definition of “urban” areas in Ireland does not include many areas at the periphery of towns and cities, urban and rural populations were distinguished by population density (table 2.5), based on the average number of inhabitants at the 1996 and 2002 census. Three categories were created for analysis, with the cut-off points (<1 person/hectare, 1-20 persons/hectare, >20 persons/hectare) chosen to give an approximately equal population in each group.

Table 2.5 Distribution of cancer cases in 1994-2003,[1] 2002 population and number of EDs, by population density tertiles

Population density	No. of cancer cases[1]	2002 population	No. of EDs
<1 person/ha	50,794	1,546,928	2,726
1-20 persons/ha	28,983	1,127,965	277
>20 person/ha	43,009	1,242,310	416

[1] all malignant cancers, excluding non-melanoma skin cancer

2.1.3.3 Socio-economic indicators

Socio-economic information for each ED was based on data from the 2002 census, which was more detailed than that contained in the 1996 census and also covered a small number of additional EDs not in the 1996 SAPS. The available variables are listed in table 1.5 and relate to: unemployment, employment type and social class, housing, car ownership, school leaving age, and elderly persons living alone. The socio-economic variables were highly correlated in time. For example, areas with high unemployment in 2002 also had high unemployment in 1996 (correlation coefficient=0.83). Similarly, the composite deprivation index (see below) was also correlated between 1996 and 2002 (correlation coefficient=0.77). The same was true when census data for 1991 were considered. This means that the choice of year should make little difference to the results. In the analysis, these socio-economic indicators (other than the composite deprivation index - see below) were categorised into quartiles based on population.

Table 2.6 Socio-economic indicators available at ED level in the 2002 census

Variable	Definition
unemployment	Proportion of the economically active population aged over 15 unemployed or seeking a first job
agricultural workers	Proportion of persons from socio-economic groups I (farmers) and J (agricultural workers)
manual workers	Proportion of persons in socio-economic groups E (manual skilled), F (semi-skilled) and G (unskilled)
non-manual/higher professional workers	Proportion of persons from socio-economic groups A (employers and managers), B (higher professional), C (lower professional) and D (non-manual)
lower social class	Proportion of persons classified as 5 to 6 on the Irish Social Class Scale[1]
early school leavers	Proportion of persons whose education ceased at, or before, age 15
overcrowded housing	Ratio of the total number of persons divided by the total number of rooms in permanent private households, therefore representing the average number of persons per room
local authority housing	Proportion of houses purchased from local authority or rented from local authority
car ownership	Proportion of persons who do not own a car
65 and older living alone	Proportion of those aged 65 and older who live alone

[1] O'Hare et al, 1991

2.1.3.4 Deprivation

The deprivation index developed by Dr Alan Kelly of the Small Area Health Research Unit was used as an index of relative deprivation at the ED level (Kelly and Teljeur, 2004). It is similar in design to the widely regarded Carstairs and Townsend indices employed in the UK (Carstairs and Morris, 1991, Phillimore et al, 1994), with certain modifications in view of differences in definition and scope between census variables in the UK and Ireland. The index is a combination of several socio-economic variables from the 2002 census, namely unemployment, social class, type of housing tenure, car ownership and overcrowding. A score was determined for each ED based on the first principal component from principal component analysis. The score was divided into quantiles, ranging from least to most deprived. Although approximate deprivation deciles are available, to provide more stable estimates the ten categories were collapsed into five, with two deciles assigned to each approximate quintile (table 2.7).

Table 2.7 Population and number of EDs included in the each deprivation category

Deprivation category	2002 population	no. of EDs	% of total population
1 (Least deprived)	926,000	684	23.6%
2	593,197	684	15.1%
3	546,843	685	14.0%
4	685,703	684	17.5%
5 (Most deprived)	1,165,460	684	29.8%

2.1.3.5 Correlations between deprivation index, population density and socio-economic variables

Table 2.8 shows the correlation coefficients between the deprivation index, population density and various individual census-based socio-economic variables. Several of the variables were highly correlated. As might be expected, population density was strongly inversely associated with the proportion of agricultural workers. The individual variables which make up the composite deprivation index were, unsurprisingly, strongly positively correlated with the overall index. There were positive correlations between the proportions of early school leavers and those classified as lower social class, the proportions unemployed and those living in local authority housing, and the proportions in local authority housing and those without a car. The percentage of elderly people living alone was not strongly correlated with any of the other variables.

2.1.3.6 Geographic distribution of deprivation index, population density and socio-economic variables

Map 2.1 shows the geographical distribution of the deprivation index. EDs which fall into the highest deprivation category are concentrated in parts of Dublin and Cork and towards the west and northwest of the country. Population density tertiles are shown in map 2.2. Only EDs in the very centre of the largest towns and cities fall into the highest tertile of population density (416 EDs; table 2.5). In most of the country, the population density is less than one person per hectare; 2,226 EDs are included in the lowest population tertile.

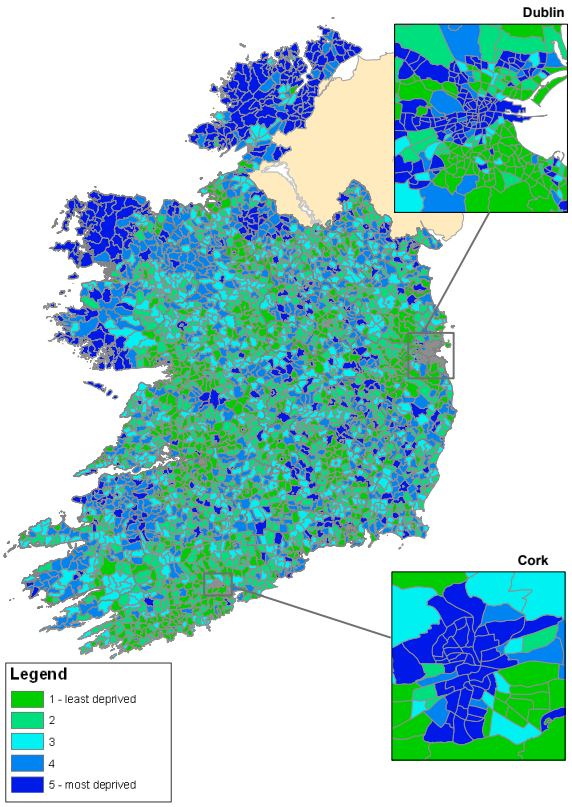
Maps 2.3-2.12 show the geographical distribution of the other census-based socio-economic variables. These were divided into 10 groups using natural breaks defined using the ArcGis function which identifies break points and maximises differences between groups (Environmental Systems Research Institute Inc., 2007). The colour ramp goes from green to blue, with areas with the lowest proportion of the variable (group 1) shown in dark green, areas with the highest proportion of the variable (group 10) shown in dark blue, and areas with intermediate values (groups 2-9) shown in a range of shades ranging from lighter green to lighter blue.

Table 2.8 Matrix of correlation coefficients for ED characteristics

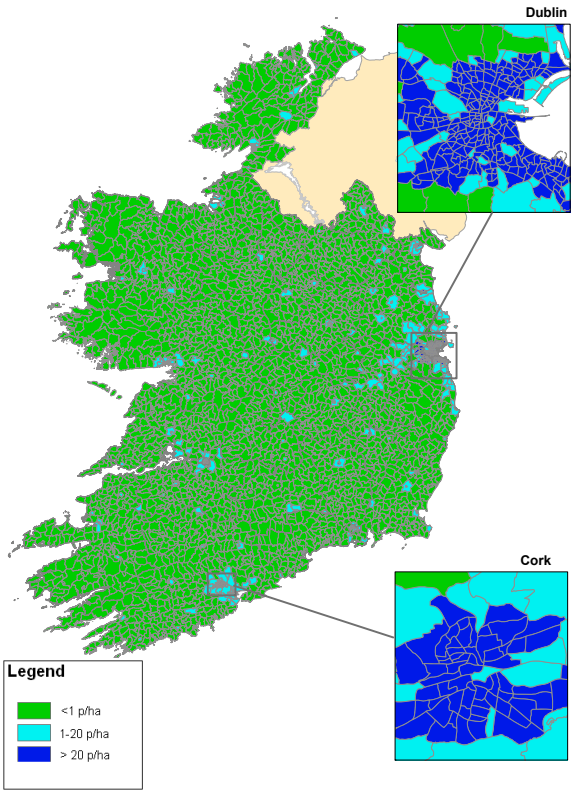
	deprivation index	population density	unemployment	lower social class	early school leaver	no car	overcrowded housing	local authority housing	% 65+ living alone	agricultural workers	non-manual worker
deprivation index	1										
population density	0.176	1									
unemployment	0.762	0.233	1								
lower social class	0.732	-0.099	0.471	1							
early school leaver	0.427	-0.275	0.265	0.531	1						
no car	0.629	0.476	0.513	0.369	0.208	1					
overcrowded housing	0.453	-0.067	0.247	0.303	0.270	0.008	1				
local authority housing	0.712	0.287	0.525	0.452	0.202	0.528	0.249	1			
% 65+ living alone	0.198	0.0002	0.143	0.175	0.167	0.357	-0.125	0.151	1		
% agricultural workers	-0.168	-0.892	-0.248	0.118	0.356	-0.443	0.069	-0.302	0.038	1	
% non-manual workers	-0.403	0.403	-0.205	-0.589	-0.676	-0.119	-0.301	-0.162	-0.157	-0.499	1
% manual workers	0.553	-0.060	0.347	0.650	0.421	0.215	0.351	0.399	0.099	0.042	-0.417

Red font=correlation >0.5; green font= correlation in range 0.4-0.5

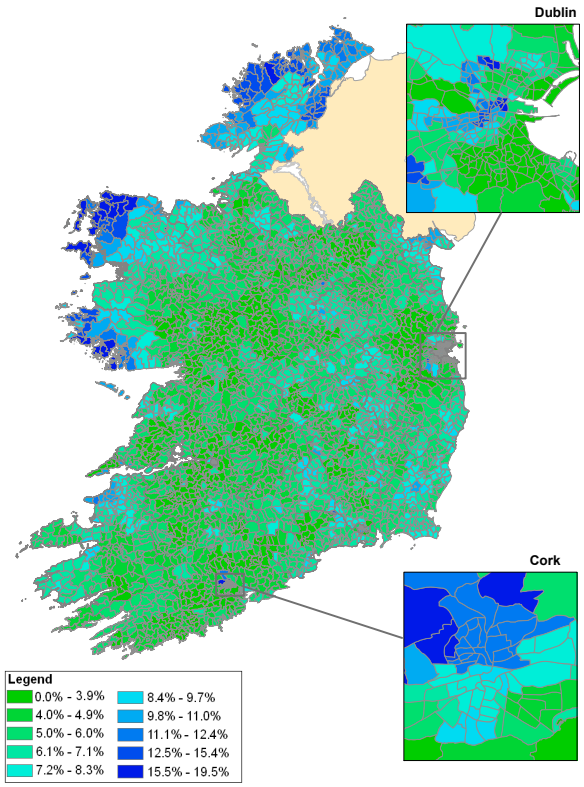
Map 2.1 Deprivation index



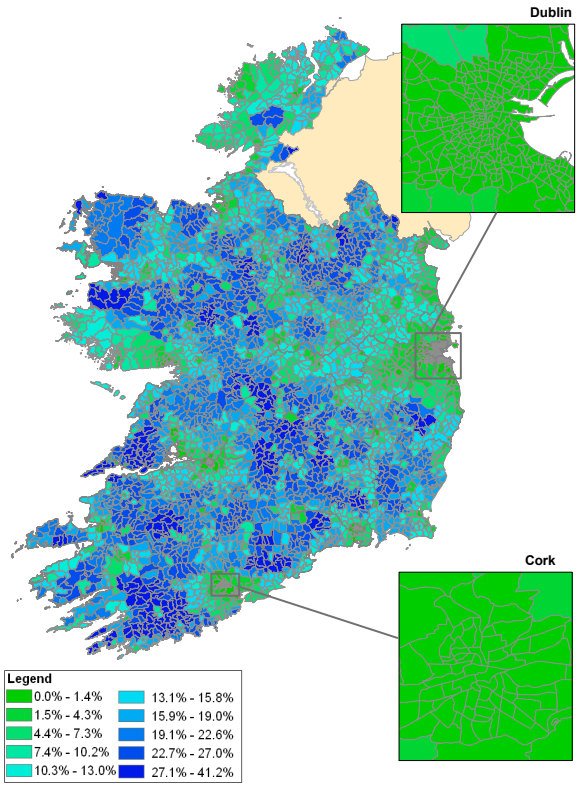
Map 2.2 Population density



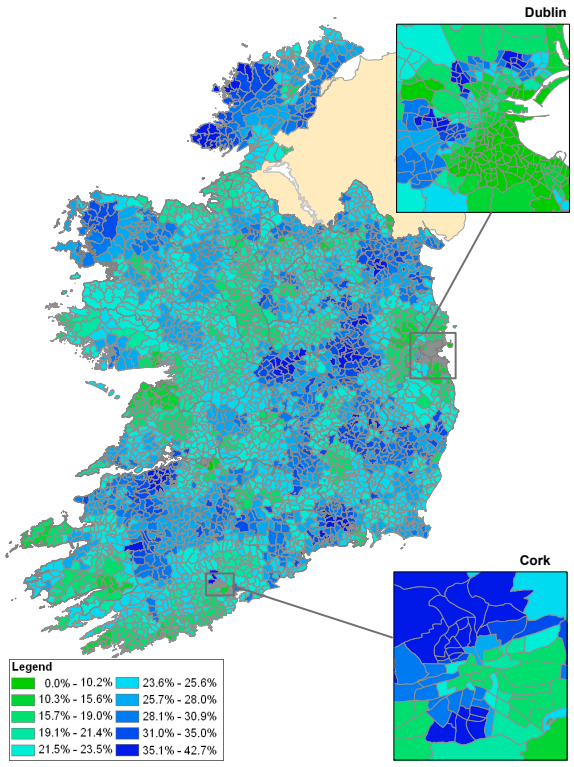
Map 2.3 Percentage unemployed



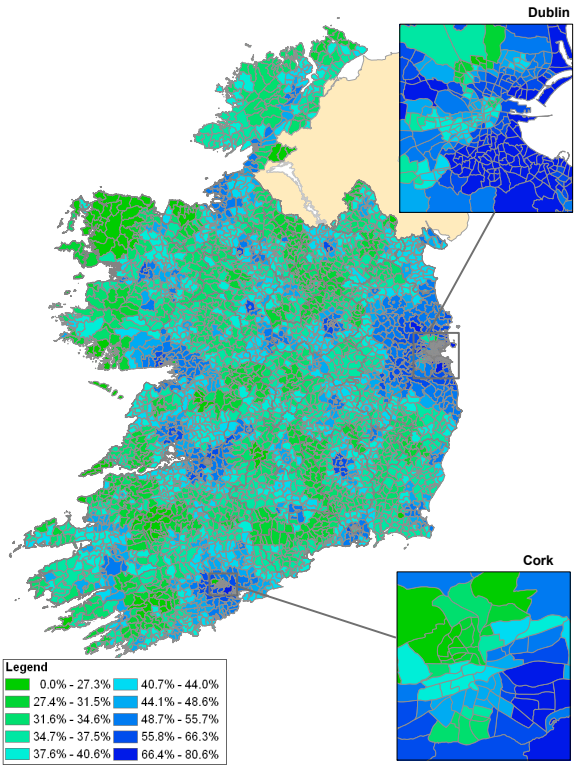
Map 2.4 Percentage of agricultural workers



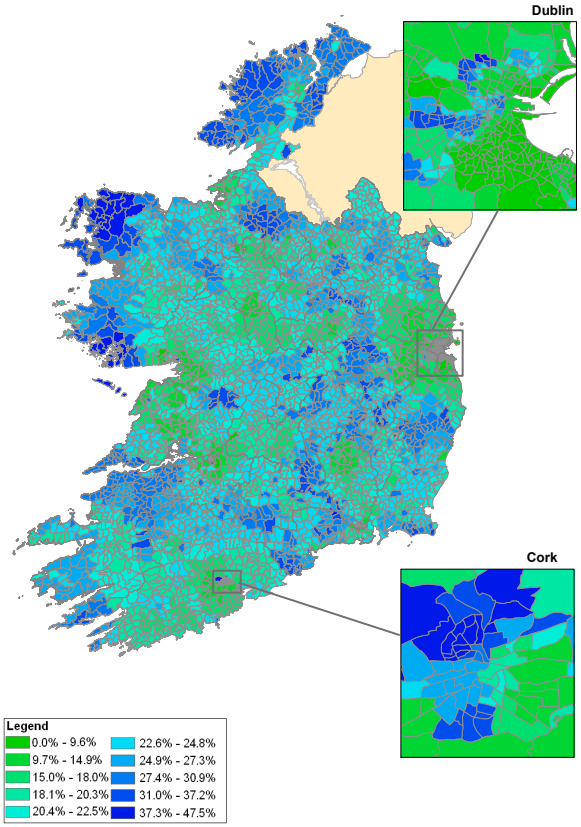
Map 2.5 Percentage of manual workers



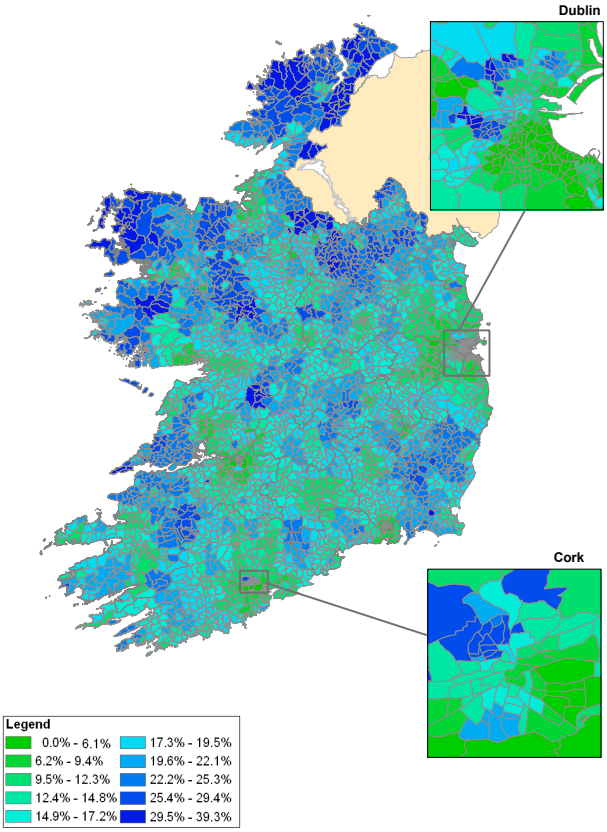
Map 2.6 Percentage of non-manual workers



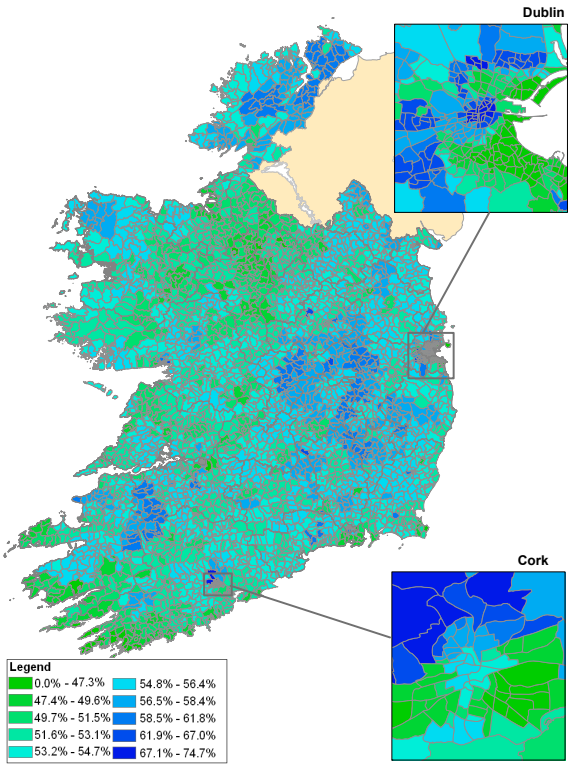
Map 2.7 Percentage in social classes 5 & 6



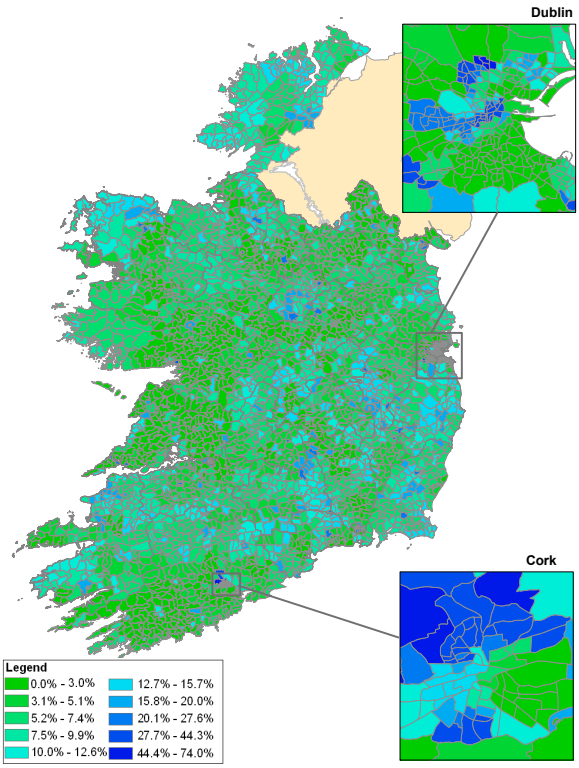
Map 2.8 Percentage of early school leavers



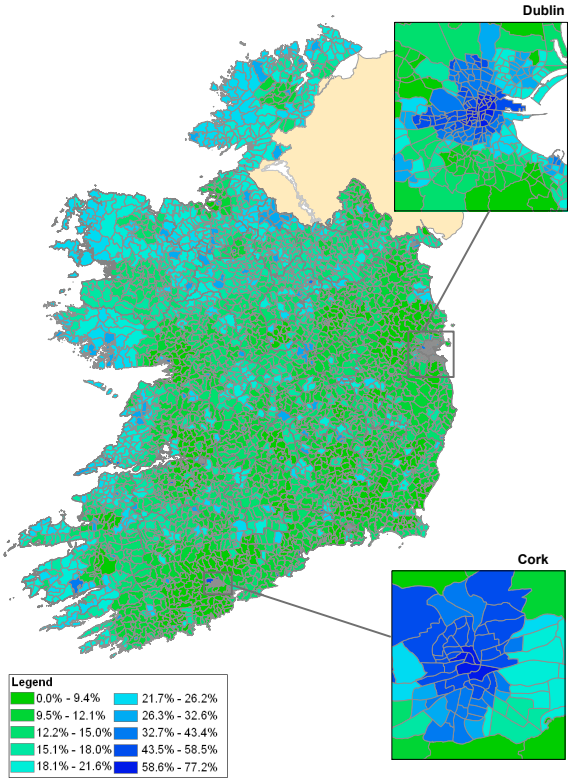
Map 2.9 Percentage in overcrowded housing



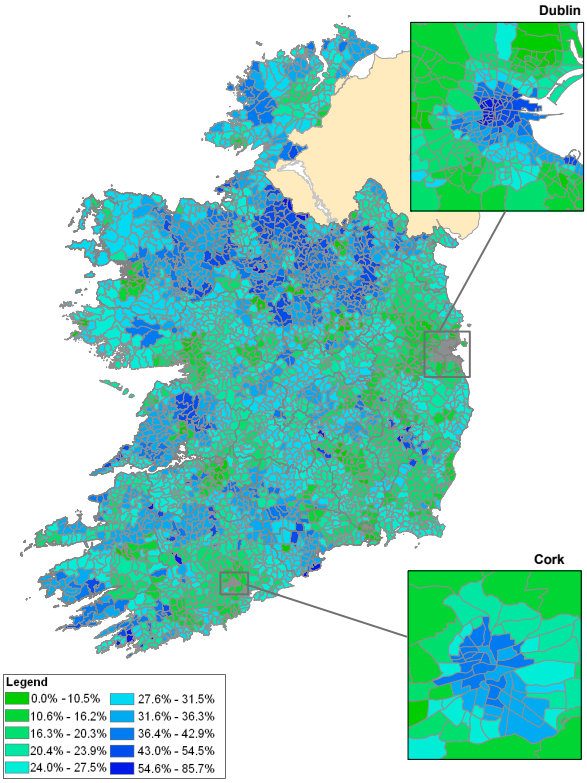
Map 2.10 Percentage in local authority housing



Map 2.11 Percentage without a car



Map 2.12 Percentage aged 65 and older living alone



2.1.4 Exposure data

The authors of the SLÁN survey (Morgan et al, 2008) provided information on various aspects of socio-economic status, diet and lifestyle. This data was collected in a population survey, conducted in 2007, which involved face-to-face interviews with more than 10,000 adults across Ireland. Although available at ED level, the information was aggregated into larger geographical areas to avoid identifying respondents. The information provided was expressed as the percentage of respondents in each geographical area, and related to the following variables:

- % in social class 6
- % in quintile five (highest) of household equivalised income
- % below 60% of median equivalised income (modified OECD equivalence scale)
- % covered by private health insurance
- % who are obese (self-reported body mass index ≥30kg/m^2)
- % with low fruit and vegetable intake (fewer than five helpings of fruit and vegetables daily)
- % with low fibre intake (less than 25g fibre daily)
- % with high intake of red and processed meat (>300g red and processed meat per week)
- % with heavy alcohol consumption (≥14 units weekly)
- % who currently smoke (daily or occasionally).

As the data was sparse, and perhaps unrepresentative at the ED level, it was not formally incorporated into the analyses in this report. Instead it is used in a purely descriptive way to add some context to the disease mapping, and to aid interpretation of the geographical patterns in disease incidence. The authors of the current report mapped the data; these maps are shown in Appendix 1.

Also shown in Appendix 1 is a map of predicted radon exposure in Ireland, derived from a report by the Radiological Protection Institute of Ireland (Fennell et al, 2002).

2.1.5 International cancer incidence data

Estimates of cancer incidence in Europe and the United States of America are taken from the GLOBOCAN 2002 software package (Ferlay et al, 2004). These estimates are sometimes quite different from the actual incidence rates given in this report for 1994-2003, for two reasons: the projections of 1999 incidence rates on which they are based may not always be accurate and they are standardised to the World, rather than the European, Standard population. However, they are useful in giving a general idea of the incidence of cancer in Ireland relative to other countries.

2.2 Statistical analysis

2.2.1 Standardised incidence ratio

In comparing cancer cases between areas or over time, two important factors must be considered - the number of people at risk and their ages. The reason and method for correcting for the number of people at risk is obvious - the number of cases is divided by the number of people resident in the area during a specified period, as reported by the census, to produce an incidence rate (or mortality rate if deaths rather than cases are being considered).

Since the risk of developing cancer risk doubles with every eight or nine years of life, an area with an older population would be expected, all else being equal, to have more incident cancer cases than an area with a younger population. There are several different approaches available to correct for age. We have used indirect standardization. This is the most appropriate method for small area comparisons, as it provides more stable rates than other standardization techniques, and works even if there is no population-at-risk in some age groups within the area (Estéve et al, 1994). For each small area *i*, we apply the national incidence rates for each age group *j* to the population counts (N) in each age group, to calculate the total expected (E) number of cancers in the area. This can be compared to the number actually found in the area, in the form of an observed (O) to expected ratio, or percentage. This is called the **standardised incidence ratio,** abbreviated to SIR**.** The SIR for any cancer for Ireland as a whole is, by definition, 1 (or 100%).

$$SIR_i = \frac{O_i}{E_i} \text{ where, } E_i = \sum_j N_{ij} \frac{O_j}{N_j}$$

2.2.2 Spatial analysis and smoothing

There are several types of geographical analysis of disease incidence:

- disease mapping, which aims to provide an estimate of the disease rate in each small area which is as close as possible to the true value;
- cluster studies, which specifically search for “clusters” - areas or groups of areas where risk is significantly higher than in the rest of the population;
- point source studies, which investigate disease risk around a "point source" of possible risk which has been defined *a priori* (e.g. an industrial site).

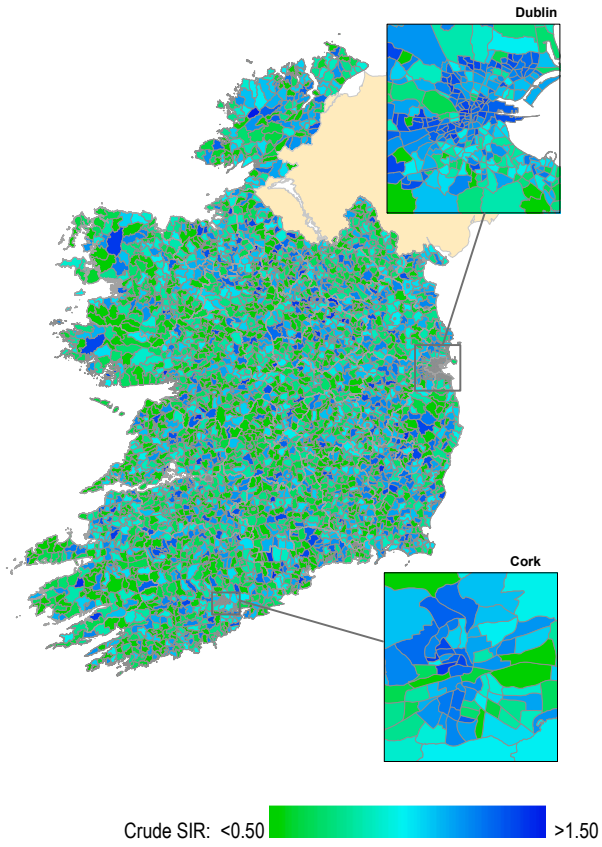
Because our primary aim was to estimate risks precisely in each small area (ED), we used disease mapping methodology.

Incidence rates, whether crude or standardised, are subject to high variability due to the small number of cases incident in each small area, and the often small population-at-risk. In many instances, areas with small populations can appear to have a particularly high or low risk, purely by chance. The average population of an ED in Ireland is about 1,145, but some are considerably smaller. One of the commonest cancers, colorectal cancer, has an incidence rate of 0.5 cases per 1,000 persons per year, so even over the 10-year period examined here, only 5 cases would be expected in a typical ED. With such small numbers, random variation is the major factor in the variation of incidence rates between EDs, and this “noise” tends to obscure any other patterns. Therefore, simply mapping the SIRs for each ED can be seriously misleading, as the SIRs tend to be more extreme in areas where the population is sparse. These areas are often the largest in area and can dominate a map visually. This is illustrated for lung cancer in map 2.13.

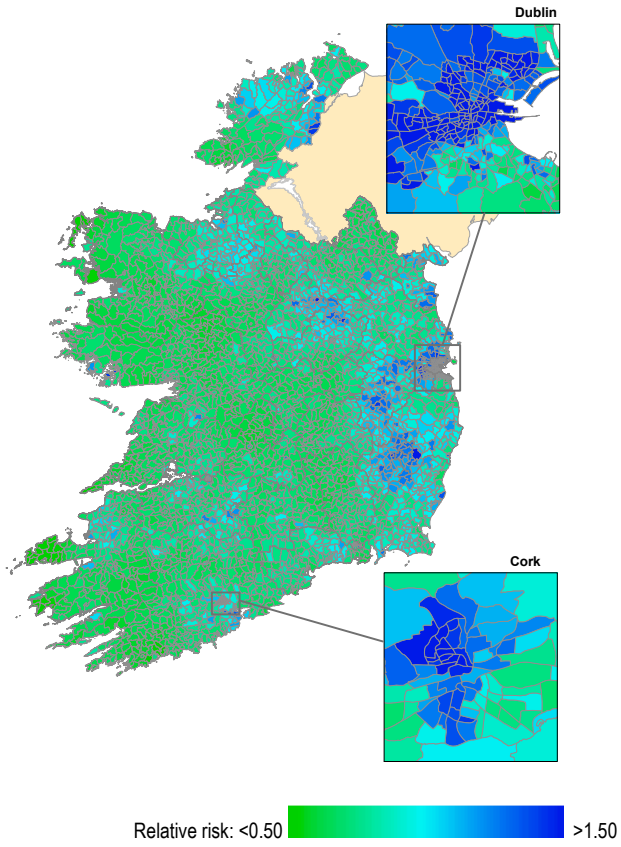
The way of dealing with this problem involves "smoothing" the estimates of disease risk (Elliott et al, 1992). Smoothing removes the noise (i.e. it smoothes out the random variation) and shows the true geographical pattern in risk more clearly. This produces relative risks (RR). The effect of smoothing is illustrated in map 2.14, which shows smoothed RRs for lung cancer, compared with the unsmoothed SIR in map 2.13.

The section below describes, in statistical terms, how the smoothed RRs were estimated. The principle of spatial smoothing is straightforward. If we assume that the risk of cancer does not vary much between areas which are close to each other, then differences between EDs are more likely to be due to random variation than to real differences in risk. The smaller the population of the area, the larger will be the element of random variation and the crude SIR will be quite an unreliable indicator of real risk. Smoothing the SIR for an ED allows us to strengthen the estimate for the ED by "borrowing strength" from adjacent areas (local smoothing) and/or from the overall/national map (global smoothing) in order to increase the stability of the estimated RR. Therefore, what smoothing does is to adjust risk estimates based on small numbers towards a local mean - based on the rates in the neighbouring areas - and also towards the national value (1.0).

Map 2.13 Lung cancer, crude SIRs: both sexes, 1994-2003



Map 2.14 Lung cancer, smoothed RRs: both sexes, 1994-2003



Many methods have been proposed for smoothing disease rates (Elliott et al, 1992). We have chosen to use a Bayesian approach (Best at al, 2005). The main advantage of Bayesian techniques is that they work well in situations of limited information and high uncertainty. They are better at accurately depicting the geographical pattern in risk than other techniques, such as non-hierarchical approaches, which are more likely to be visually misleading (Pascutto et al, 2000).

The SIRs were smoothed by estimating relative risks using **conditional autoregressive models** (CAR) (Clayton and Kaldor, 1987) based on a spatial Poisson model with two random effects, as follows:

$$Oi \sim Poisson(E_i\theta_i)$$
$$\log(\theta_i) = \alpha + h_i + b_i$$

where

O_i is the observed number of cancer cases in area *i;*

E_i is the expected number based on national incidence rates;

b_i is a spatially structured random effect (which is given a CAR prior distribution);

h_i is a random effect which models the unstructured heterogeneity;

α is the intercept; and

θ_i is the estimated relative risk.

Use of CAR models is widespread in disease mapping and this particular model is known to be appropriate in most situations (Lawson et al, 2000, Best et al, 2005). Other methods (e.g. kernel smoothers, mixture models) seem to give poorer results than CAR (Lawson et al, 2000). Although risk estimates can be somewhat underestimated, CAR models have a high specificity (Richardson et al, 2004), and this conservative approach means that high or low estimates are more likely to be real. However, with this method, as with any smoothing method, it is possible that areas of genuinely high risk may be missed by smoothing with neighbouring areas. The method also assumes that risk varies smoothly at the scale studied, an assumption which may not be justified if environmental effects at a purely local level (e.g. air pollution) are important.

We fitted our models using Markov Chain Monte Carlo (MCMC) algorithms with WinBUGS software (Lunn et al, 2000). Estimates were checked to ensure convergence had been reached. A burn-in of 50,000 iterations (or more if convergence was not reached) was performed for each model and the posterior distributions were derived using one in three iterations from the subsequent 10,000 iterations.

Relative risks (RR) were mapped for each cancer site individually using ArcMap 9.2. For those cancers which affect both sexes, maps are included for the sexes combined and for males and females separately. County boundaries are shown faintly on the maps to help the reader with geographical orientation; a map of the counties is contained in Appendix 4. To facilitate comparisons between cancer sites, each map is shown using the same colour ramp, which goes from dark green for an estimated RR less than 0.50 to dark blue for a RR higher than 1.50 (i.e. the same colour represents the same value of RR on each map). Appendix 3 contains summary information from the mapping of each cancer site, including average numbers of cases per ED, and mean crude SIR and smoothed RRs.

2.2.3 Poisson regression: ED characteristics and cancer incidence

We used Poisson regression to investigate the relationship between the risk of cancer and deprivation, population density and other area-based socio-economic variables. The number of new cancer cases in ED i, age group j, is assumed to be Poisson distributed. Fitting the model produces an estimate of the risk of cancer for each quantile

of the explanatory/independent variable(s), relative to a common reference category (e.g. quartile with lowest unemployment) - that is, it produces a RR.

The analysis proceeded as follows for each cancer site separately. The first analysis related to deprivation index. The least deprived quintile (deprivation group 1) was taken as the reference category and relative risks were computed for areas in deprivation categories 2-5. The risk estimates were adjusted for population density, since this is an important confounder of the relationship between deprivation and cancer incidence (see maps 2.1 and 2.2). In the second analysis, we built multivariate models using population density and the other variables shown in table 2.5 as candidate explanatory variables. Since some of the variables were highly correlated (table 2.6), their inclusion in the same model was not appropriate. To deal with this, we first created a multivariate model where all of the variables, except those relating to occupational group, were considered for inclusion (i.e. population density, unemployment, lower social class, overcrowded housing, local authority housing, car ownership and 65 and older living alone). We retained in the final models those variables which provided the best fit to the data, as assessed by likelihood ratio tests and the Akaike information criterion (AIC). We then built a model exploring the relationship between occupational group (percentages of agricultural workers, manual workers and non-manual workers) and cancer. These models included the same adjustment factors as the previous multivariate model, except that population density was not included (since population density and percentage of agricultural workers was so highly correlated). We presented results for the occupational group with the best model fit. For those cancers which affect both sexes, the models were created using data for both males and females. The results of these analyses are contained in the individual chapters relating to each cancer site. In addition, Appendix 2 includes summary tables which provide an overview of the results.

Using Poisson regression to model relative risk based on small-area has limitations. In particular, there may be overdispersion, which occurs when the observed variance is higher than expected (Breslow, 1984). This is because Poisson models do not have a dispersion parameter and the geographical distribution of the data makes it likely that dispersion will be high. In practice, the relative risks will be correctly estimated but the confidence intervals may be under-estimated.