A point process model for the spatial distribution of brain tumours in relation to exposure from mobile phones

(Short: Point process model for brain tumours and mobile phones)

Kathrine Grell[1,2], Peter J. Diggle[3], Kirsten Frederiksen[1], Joachim Schüz[4], Elisabeth Cardis[5], and Per Kragh Andersen[2]

[1]Unit of Statistics, Bioinformatics and Registers, Danish Cancer Society Research Center, Copenhagen, Denmark

[2]Section of Biostatistics, Department of Public Health, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

[3]CHICAS, Faculty of Health and Medicine, University of Lancaster, Lancaster, UK

[4]Section of Environment and Radiation, International Agency for Research on Cancer (IARC), Lyon, France

[5]Centre for Research in Environmental Epidemiologi (CREAL), Barcelona, Spain

April 28, 2014

Abstract (max 250 words in SiM)

The spatial distribution of the tumours is relevant to take into account when investigating the relation between brain tumours and the exposure from radio frequency electromagnetic fields caused by mobile phone use. This issue resembles investigating spatial aggregation of a disease around a source of potential hazard which can be analysed using point process modelling. We propose to use a Poisson point process model similar to one used in environmental epidemiology when including tumour localisation in modelling the association between brain tumours and mobile phones. The spatial distribution is then a distribution over a sample of patients rather than over multiple disease cases within one geographical area. We show how the distance relation between tumour and phone can be modelled nonparametrically and with various parametric functions, how covariates can be included in the model and how to test for effect of the distance. The models are applied to a subset of the data from the Interphone Study, a large multinational case-control study on the association between brain tumours and mobile phones, where we see that the impact of distance between tumour and preferred ear for mobile phone use on the tumour intensity is significant.

Keywords (max 6 in SiM)

Poisson point process; spatial point pattern; brain tumours; mobile phones; radiofrequency fields

1 Introduction

Analysis of spatial point patterns in geographical or spatial epidemiology can be used to investigate spatial aggregation of a certain disease in addition to what is due to environmental heterogeneity. This is useful also in environmental epidemiology when examining a possible raised incidence of cases of a disease around a source of potential environmental hazard. Diggle [1] proposed a point process modelling approach to test for possible clusters in the vicinity of fixed locations motivated by a concern for an elevated risk of certain types of cancers near a former industrial waste incinerator[1] and nuclear installations[2]. This concern for clustering of cancer cases near nuclear installations and waste incinerators had increased in the preceding years[3–6], and in 1989 the Royal Statistical Society held a meeting on 'cancer near nuclear installations'[7] where Diggle's point process model were one of the methodological contributions. The first version of the point process model suggested estimating the background or baseline intensity[1], while a newer version introduced a conditional approach that eliminates the baseline intensity from the estimation[8, 9].

We have adapted this point process model to a setting where the area of interest is a part of the human body instead of a geographical area, each person contributes with the localisation of their disease as only one point within this area, and the putative source of exposure is a device external to but held near to the body. Hence, we investigate the spatial distribution of tumour localisations over a sample of persons, and not multiple tumours within one person. The latter would correspond to the models for cancer cases within a geographical area. Our setting also differs from the geographical setting since it concerns

3-dimensional modelling.

Our work was motivated by a wish to use information on brain tumour localisation when modelling the association between brain tumours and exposure from radio frequency electromagnetic fields (RF-EMF) emitted by mobile phones. Localisation is crucial to take into account when investigating this association, since the absorption of energy from RF-EMF in human tissue depends greatly on the distance from the radiation source. Consequently, studies have dichotomised the distance between the mobile phone and the brain tumour for use in their analyses[10, 11] or calculated a specific exposure measure related to the energy absorption for each tumour[12, 13]. We propose in this paper a, for this setting, new statistical model including the localisations. Our specific aim is to introduce an analyses plan for investigating the 3-dimensional spatial distribution of disease points in relation to source points, starting with an exploratory analysis and subsequently using the point process model together with various regression models for the distance relation between disease points and point source.

The multinational case-control study Interphone on the association between brain tumours and mobile phone use recorded the specific localisations of the tumours in a 3-dimensional grid map of a standard human brain[14]. The study comprises detailed information on past mobile phone use obtained via interviews and diagnostic information from hospital records from the 13 participating countries (Australia, Canada, Denmark, Finland, France, Germany, Israel, Italy, Japan, New Zealand, Norway, Sweden, and the UK), and we have used it as an illustrative example to show how an exploratory analysis and the point process model can be applied.

This paper is organized as follows. In Section 2 we present our motivating data from the Interphone Study. We describe the exploratory analysis and show the results from this when applied to the Interphone data in Section 3. In Section 4 we formally describe the point process model, different models for the distance relation, and how to do inference, together with the relating results. Finally a brief discussion follows in Section 5.

2 Motivating data

The Interphone Study has been described in detail elsewhere[14]. Briefly, cases were between 30 and 59 years of age, when diagnosed with a first primary glioma, meningioma or acoustic neurinoma during study periods of 2–4 years between 2000 and 2004. All cases were histologically confirmed or based on unequivocal diagnostic imaging. Gliomas grow from glial cells and can therefore occur all over the brain, whereas meningiomas and acoustic neurinomas are restricted to certain areas of the brain since they grow from respectively meningeal cells, the layer of tissue covering the brain and spinal cord, and the nerve sheath cells covering the acoustic nerve. Even though gliomas are not restricted to parts of the brain, the spatial distribution is not completely random but there is an unknown spatial baseline distribution[15, 16].

Considering only gliomas, since their origin in the brain is not spatially restricted, the Interphone data comprise 2710 glioma cases and 2974 matched controls. Our model is based only on the cases and hence removing any differential bias between cases and controls both with regard to recall errors and selection of participants[17, 18]. Of the 2710

cases the specific localisation information is registered for 1530 cases. Tumour localisations were recorded by neuroradiologists in each study centre on a 3-dimensional grid map of the human brain made up of 1 cm cubes (voxels) in the program GridMaster[19] made specifically for the Interphone study. These recordings were based on radiological images, MRI and CT scans, when available and otherwise on radiology reports. The neuroradiologists, blind to the data on mobile phone use, were also asked to record their best estimate of the tumour origin. For 906 tumours only one voxel was marked as the origin and for the remaining 624 tumours either none or several voxels were marked. The tumours with none or several voxel origins are discarded from the analyses in this paper, since it is not straightforward how to define their origin as a single point.

Detailed information on past mobile phone use was collected by interview with study subjects or proxies, and this included use of hands-free device, preferred side of the head for mobile phone use and number of calls and call time. Of the 906 glioma cases with one voxel origin 560 were regular phone users, defined at that time as a person that had at least one call per week for a period of 6 months or more, and the 346 non-regular phone users were defined as not exposed. Consequently the non-regular users are not included in our analyses. The lifetime cumulated call time and number of calls were calculated for all regular users as part of the original Interphone study using an imputation procedure for persons with missing responses to questions about mobile phone use[14]. The cumulative use excluded use of mobile phones with hands-free devices[20], since this reduces the amount of exposure to the head[21]. Few of the regular phone users had a cumulated call time of 0 hours when accounting for hands-free devices. The interview also had a question about

which side of the head was generally used for mobile phone use and of the 560 regular phone users 171 reported left side, 307 reported right side, 60 reported both, and for the remaining 22 regular phone users the information was missing. We will refer to the persons preferring the right side of the head as 'right-users', the persons preferring the left side of the head as 'left-users', and accordingly persons preferring both sides as 'both-users'. We include only the tumours from persons preferring one side in our analyses, because it is less clear how to define the source point for the both-users.

The exposure from mobile phones can be measured as the specific absorption rate of energy (SAR), which is a measure of the energy absorbed in the body per unit mass of tissue, and the spatial distribution of SAR within the head depends on different parameters such as the position of the phone in relation to the head and the frequency band[22]. Even though it is a simplification, we assume that the energy is emitted at the ear on the side of the head where the phone was reported generally used. The ear canals were identified to be fully contained within 48 voxels on each side of the GridMaster head and we have defined the geometric midpoint of the outer area of these as the ear.

In total we have 478 glioma cases with localisation data where only one voxel were marked as origin of the tumour and of which 171 reported left-side use and 307 reported right-side use. We will refer to the midpoint of the origin voxels as the tumour points. The spatial reference for the tumour origins is hence the coordinates of the tumour points, and the source point is either the coordinates for the left ear or the right ear.

3 Exploratory analysis

Each tumour is identified with a single reference location $\mathbf{x} = (x_1, x_2, x_3)$ chosen as the midpoint of the voxel marked as the tumour origin. The two ears are similarly identified with locations $\mathbf{x}_L$ and $\mathbf{x}_R$ as described in the previous section. For left-users or right-users, only $\mathbf{x}_L$ or $\mathbf{x}_R$, respectively, enters the model for the spatial distribution of $\mathbf{x}$. The preliminary exploratory analysis we describe here is nonparametric estimation of the density distributions for the left-users' tumour points and the right-users' tumour points. If the two distributions are different in the direction of the ear, it supports the hypothesis about the distance from preferred ear being important for tumour localisation.

Let the probability density for the left-users' tumour points $\mathbf{x}$ be a product of a baseline density f_0, a normalising constant $c(\theta)$ and a function h that shifts the density towards the left ear $\mathbf{x}_L$ as a function of the distance $\mathbf{x} - \mathbf{x}_L$. The spatial distribution of RF energy in the brain is similar for the left half and the right half of the brain[22], hence we assume h to be the same for the left-users and the right-users. We write the two probability densities as in (1).

$$f_L(\mathbf{x}) = c(\theta) f_0(\mathbf{x}) h(\mathbf{x} - \mathbf{x}_L; \theta) \quad \text{and} \quad f_R(\mathbf{x}) = c(\theta) f_0(\mathbf{x}) h(\mathbf{x} - \mathbf{x}_R; \theta) \tag{1}$$

We used the R-package ks[23] for computing the kernel estimators with a Gaussian kernel and a multivariate smoothed cross-validated bandwidth selector, but it is important to note that the nonparametric density estimates can be sensitive to the choice of bandwidth selector[24]. In Figure 1 are shown the 3-dimensional kernel density estimates as a series of nested contours, where the contours are the boundaries of the highest density regions[25].

We see that the density for the left-users is skewed towards the left ear and the density for the right-users is skewed towards the right ear, clearly indicating that the two densities differ. A test of the hypothesis $f_L = f_R$ using the integrated L_2-error with the optimal plug-in bandwidth selector for functional estimators, also implemented in ks, gives however, $p = 0.29$ indicating that the densities are not significantly different. Since this might reflect that the sample size is quite small for nonparametric estimation we find it, based on the skewness in Figure 1, sensible to continue with the point process model.

4 The point process model

The incidence of tumours, still assumed to be 3-dimensional points, in a human brain can be regarded as a realisation of a spatio-temporal point process. Not allowing for multiple tumours, the interest here is only on the first realisation of the process in each person. These first realisations are independent of each other and together they form an inhomogenouos Poisson process on the volume of the brain A with a certain intensity denoted $\lambda_0(\mathbf{x})$ for persons not using a mobile phone.

Following Diggle and Rowlingson [8] we consider a model where the left-users and the right-users form independent Poisson processes with respective intensities $\lambda_L(\mathbf{x})$ and $\lambda_R(\mathbf{x})$. We write these intensities as a product of a baseline intensity $\lambda_0(\mathbf{x})$ and a function that depends on the distance from the point to the source point, that is the distance from tumour to the left ear for the left-users and to the right ear for the right-users, and covariates z. The intensities are given by (2) and (3), where ρ is a nuisance parameter that

relates to the relative number of left- and right-users.

$$\lambda_L(\mathbf{x}) = \lambda_0(\mathbf{x})g(\mathbf{x} - \mathbf{x}_L, z; \theta) \tag{2}$$

$$\lambda_R(\mathbf{x}) = \rho\lambda_0(\mathbf{x})g(\mathbf{x} - \mathbf{x}_R, z; \theta) \tag{3}$$

We still assume the spatial distribution of RF energy in the two brain halves to be symmetrical by having the same distance relation g in the two intensities. Different forms of the distance relation will be discussed in Section 4.1.

The superposition of the two Poisson processes is also a Poisson process with intensity $\lambda_L(\mathbf{x}) + \lambda_R(\mathbf{x})$. In this superposition we define a binary random variable Y with $Y_i = 1$ if the ith event in the superposition is from the process of the left-users and $Y_i = 0$ if it is from the process of right-users. Hence all events are labelled either left-user or right-user. Conditional on the locations the labels are mutually independent, and consequently we now have a binary regression model with spatially dependent probabilities defined as the conditional probability of seeing a left-user point given the location $\mathbf{x}$ as in (4).

$$p(Y = 1|\mathbf{X} = \mathbf{x}) = \frac{\lambda_L(\mathbf{x})}{\lambda_L(\mathbf{x}) + \lambda_R(\mathbf{x})} = \frac{g(\mathbf{x} - \mathbf{x}_L, z; \theta)}{g(\mathbf{x} - \mathbf{x}_L, z; \theta) + \rho g(\mathbf{x} - \mathbf{x}_R, z; \theta)} \tag{4}$$

This model has several advantages, since it eliminates the nuisance function $\lambda_0(\mathbf{x})$ representing the spatial baseline distribution of the tumours, and provides us with an easy framework to do inference and introduce covariates z. It should be noted that covariates introduced here only influence g and not the baseline intensity $\lambda_0(\mathbf{x})$.

Since the labels 'right-user' and 'left-user' given the locations are then independent Bernoulli trials it is straightforward to write down the loglikelihood for ρ and θ. Let

$\mathbf{x}_1, \ldots, \mathbf{x}_{n_L}$ be the n_L left-users and $\mathbf{x}_{n_L+1}, \ldots, \mathbf{x}_{n_L+n_R}$ the n_R right-users, then the loglikelihood is defined as in (5).

$$\begin{aligned}
\log L(\theta, \rho) &= \sum_{i=1}^{n_L} \log(p(Y_i = 1|\mathbf{X} = \mathbf{x}_i)) + \sum_{i=n_L+1}^{n_L+n_R} \log(1 - p(Y_i = 1|\mathbf{X} = \mathbf{x}_i)) \\
&= \sum_{i=1}^{n_L} \log(g(\mathbf{x}_i - \mathbf{x}_L, z_i; \theta)) + \sum_{i=n_L+1}^{n_L+n_R} \log(\rho g(\mathbf{x}_i - \mathbf{x}_R, z_i; \theta)) \\
&\quad - \sum_{i=1}^{n_L+n_R} \log(g(\mathbf{x}_i - \mathbf{x}_L, z_i; \theta) + \rho g(\mathbf{x}_i - \mathbf{x}_R, z_i; \theta)) \qquad (5)
\end{aligned}$$

When $g = 1$ the loglikelihood reduces to

$$\log L_0(\rho) = n_R \log \rho - (n_L + n_R) \log(1 + \rho)$$

with maximum at $\hat{\rho}_0 = n_R/n_L$. Otherwise the parameter estimates $\hat{\rho}$ and $\hat{\theta}$ are obtained by numerical maximization of (5). Model-based standard errors can be obtained by evaluation of the observed information matrix if these are proven reliable, otherwise Monte Carlo standard errors can be computed by bootstrapping. Tests of hypotheses about θ can be accomplished by the usual likelihood ratio test comparing the test statistic $D = 2(\log L(\hat{\theta}, \hat{\rho}) - \log L(\theta_0, \hat{\rho}_0))$ with critical values from the χ^2-distribution with degrees of freedom equal to the number of parameters constrained. However, consistency between the χ^2-distribution and the empirical null distribution should be examined. The empirical null distribution of the likelihood ratio test statistic can be found via a simulation study by simulating left- and right-marks for the existing tumour points based on the conditional probability in (4) using $\hat{\rho}_0$ from the original dataset. The χ^2-approximation to the null distribution can then be evaluated by plotting the observed quantiles from the empirical null distribution against the quantiles from the relevant χ^2-distribution. If the Q-Q plot

does not follow the diagonal, the χ^2-approximation is questionable, and the empirical p-value calculated from the simulation should be used. The empirical p-value is calculated as $p = (r + 1)/(n_{sim} + 1)$, where r is number of simulated test statistics larger than or equal to the observed and n_{sim} is the number of simulations.

4.1 Specification of the distance relation

The distance relation between tumour points and ear points is described via the function g, and therefore the functional form of g should depend on knowledge about how the energy from RF-EMF is distributed within the human brain. The obvious form is a continuous decreasing function with $g \to 1$ for $\mathbf{x} - \mathbf{x}_L \to \infty$ or $\mathbf{x} - \mathbf{x}_R \to \infty$, so that the impact from the phone is highest where the phone is held and disappears when the distance to the phone grows large. In fact, the absorption rate of RF energy is declining through the brain[26]. Furthermore g should be parameterised such that if the intensity is independent of the distance between tumour and ear then $g = 1$ for all $\mathbf{x}$.

First, however, it is useful to introduce a semiparametric model to visualise g and suggest suitable parametric forms. Using the framework of generalised additive modelling as in Rodrigues et al. [27] we have $\text{logit}(P(Y = 1|\mathbf{x})) = \log\left(g(\mathbf{x} - \mathbf{x}_L)/\rho g(\mathbf{x} - \mathbf{x}_R)\right)$ which we modelled by a smooth nonparametric function depending on the distance to the left ear for all users $s(\mathbf{x} - \mathbf{x}_L)$, and $\text{logit}(P(Y = 0|\mathbf{x})) = \log\left(\rho g(\mathbf{x} - \mathbf{x}_R)/g(\mathbf{x} - \mathbf{x}_L)\right)$ which we modelled by a smooth nonparametric function depending on the distance to the right ear for all users $t(\mathbf{x} - \mathbf{x}_R)$. We used the function gam in the R-package mgcv[28] with a

penalised regression spline, where the smoothness is estimated as part of the fitting, to fit s and t. The two smoothing functions are shown in Figure 2 together with pointwise 95% confidence limits. The vertical marks on the x-axis indicate the actual distances in the data explaining why the confidence bands are very large for distances below 50 mm. From the figure we conclude that a decreasing function is justified and that we, due to the similarity between the fitted s and t, can use the same function to model the distance relation to the left ear and to the right ear.

Turning to parametric specifications of g the safe approach is to start with a simple piecewise constant function, analogous to plotting the impact from the phone as a histogram over distance between tumour and ear. We defined this function as in (6) where the change points were defined using the actual distances in the data to make sure there were sufficiently many cases within each interval to estimate the value of the function. When the discontinuities are known the model corresponds to a logistic regression model. Notice that for the 'standard brain' used here there is no brain tissue within 15 mm from the ear, and the midline of the brain is 85 mm from the ear.

$$g(\mathbf{x}-\mathbf{x}_L;\boldsymbol{\alpha}) = \begin{cases} \alpha_1 & if \quad \mathbf{x}-\mathbf{x}_L \in (0;55] \\ \alpha_2 & if \quad \mathbf{x}-\mathbf{x}_L \in (55;75] \\ \alpha_3 & if \quad \mathbf{x}-\mathbf{x}_L \in (75;95] \\ \alpha_4 & if \quad \mathbf{x}-\mathbf{x}_L \in (95;115] \\ 1 & if \quad \mathbf{x}-\mathbf{x}_L > 115 \end{cases} \tag{6}$$

The value of g is fixed at 1 for the last interval, since g can only be determined up to a constant, and in agreement with there being no effect of the phone after a certain distance.

To ensure that the distance relation is a decreasing function we can add the constraint $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4 \geq 1$ to (6). The maximum likelihood estimates and the Monte Carlo standard errors for this model both without and with the decreasing restriction are shown in Table 1 and the functions are shown in Figure 3. The estimate of the nuisance parameter ρ is not shown for any of the models we present, but is each time estimated close to n_R/n_L as expected. We have reported the Monte Carlo standard errors from bootstrapping 1000 times and not the model-based standard errors, since simulations (not shown) suggested a poor consistency between the two. Even though the function starts with an increase from the first to the second interval, there is such a large uncertainty around the first interval due to few cases, that it does not contradict the hypothesis about the distance relation being decreasing, especially when also taking the nonparametric modelling shown in Figure 2 into account. Testing $g = 1$ results in the likelihood ratio test statistic $D = 21.8$ for the model with the piecewise constant function and $D = 20.5$ when the decreasing restriction is added. Simulation studies (not shown) suggested that the consistency betwen the null distribution and the χ^2_4-distribution was poor and therefore we calculated the empiral p-values based on 1000 simulations to be $p = 0.00100$ for both, hence the hypothesis of no point source effect is rejected for both models. We tried also to model g with different change points and smaller intervals which gave similar results (not shown).

The spatial distribution of RF energy in the human brain is continuous and the estimates for the piecewise constant and decreasing g-function can be used to decide on a continuous form for g. Diggle [2] suggested the function $g(\mathbf{u}; \alpha, \beta) = 1 + \alpha \exp(-\beta \mathbf{u}'\mathbf{u})$,

but based on the results from the previous g-functions we chose the continuous form

$$g(\mathbf{x}-\mathbf{x}_L;\psi,d_0,d_1)=1+\psi\begin{cases}1 & \text{if} \quad \mathbf{x}-\mathbf{x}_L\in(0;d_0]\\ 1-\left(\frac{1}{d_1-d_0}\right)(\mathbf{x}-\mathbf{x}_L-d_0) & \text{if} \quad \mathbf{x}-\mathbf{x}_L\in(d_0;d_1]\\ 0 & \text{if} \quad \mathbf{x}-\mathbf{x}_L>d_1\end{cases} \tag{7}$$

where $\psi\geq 0$ and $0<d_0<d_1$. The function (7) is constant equal to $1+\psi$ until the distance from the ear reaches d_0 and then it declines linearly until it equals 1 from d_1 and onwards. The estimate of d_0 has to be close to the boundary of the brain at 15 mm for it to be in agreement with the knowledge about spatial distribution of SAR. We maximised the loglikelihood over a grid of (d_0,d_1)-values and the grid pair with the highest loglikelihood value and corresponding estimate of ψ are shown in Table 2 together with the bootstrapped standard error. The grid values were chosen such that there were at least both one left-user and one right-user with distance to preferred ear in the interval $(0;d_0]$ and at least one of both in the interval $(d_1;\infty)$, consequently, our (d_0,d_1)-values were ranging from 45 mm to 140 mm in steps of 5 with the minimum distance between the two d-values being 10 mm. The g-function is shown together with the two previous in Figure 3. We do not obtain usual standard errors for the parameters d_0 and d_1 when estimating the loglikelihood over a grid of (d_0,d_1)-values, however, we would still like to know how well the estimates for these parameters are determined. A simple way to do this is to plot the maximum loglikelihood values for each pair in the (d_0,d_1)-grid and see whether this is a monotone function of the two parameters with a unique maximum. Figure 4 shows the maximum likelihood value for all pairs of (d_0,d_1) with the red square showing the maximum value. The maximum is unique, but the estimates $\hat{d}_0=90$ and $\hat{d}_1=100$ imply that the level of energy absorption

is constant until 90 mm away from the phone, which in the horisontal plane is in the opposite brain half, contradicting that almost all energy is absorbed within the nearest half[22, 29]. The relevant hypothesis to test is $\hat{\psi} = 0$, since this corresponds to no effect. The likelihood ratio test statistic for this hypothesis was $D = 20.9$, and since the χ^2_1-distribution approximated the null distribution poorly we calculated the empirical p-value based on 1000 simulations, of which less than 1% did not converge, to be $p = 0.00101$. This is in agreement with what we have seen for the previous models, but because of the high estimates of d_0 and d_1 this model is biologically implausible.

4.2 Adding covariates

Covariates influencing the distance relation can easily be added to the model. As an example an obvious covariate to include in our case is phone use z_{ph}. Lifetime cumulative mobile phone use measured as call time accounting for hands-free devices was reported as a continuous covariate. The unit is 1000 hours and since the variable is skewed we have taken the 4th root of it. Figure 5 shows distance to the preferred ear for left-users and right-users plotted against the transformed call time together with the best linear fit and a smoothed polynomial fit. There is no clear trend in the figures and therefore no support to the hypothesis that cumulative call time influences the distance relation. However, we will add cumulative call time as a covariate to our model, since the only observed association in the main Interphone article was for this exposure metric[20]. We would expect the level of phone use to increase the intensity no matter how large the distance between tumour

and phone is. Therefore we include it in the piecewise constant and decreasing g as in (8) with the restrictions $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4 \geq 1$ and $\phi \geq 0$ or in the continuous g as in (9) with $\psi, \phi \geq 0$ and $0 < d_0 < d_1$.

$$g(\mathbf{x}-\mathbf{x}_L, z_{ph}; \boldsymbol{\alpha}, \phi) = \begin{cases} (1+\phi z_{ph}) \cdot \alpha_1 & if \quad \mathbf{x}-\mathbf{x}_L \in (0;55] \\ (1+\phi z_{ph}) \cdot \alpha_2 & if \quad \mathbf{x}-\mathbf{x}_L \in (55;75] \\ (1+\phi z_{ph}) \cdot \alpha_3 & if \quad \mathbf{x}-\mathbf{x}_L \in (75;95] \\ (1+\phi z_{ph}) \cdot \alpha_4 & if \quad \mathbf{x}-\mathbf{x}_L \in (95;115] \\ 1 & if \quad \mathbf{x}-\mathbf{x}_L > 115 \end{cases} \tag{8}$$

$$g(\mathbf{x}-\mathbf{x}_L, z_{ph}; \psi, \phi, d_0, d_1) = 1 + (\psi + \phi z_{ph}) \cdot \begin{cases} 1 & \text{if} \quad \mathbf{x}-\mathbf{x}_L \in (0; d_0] \\ 1 - \left(\frac{1}{d_1-d_0}\right)(\mathbf{x}-\mathbf{x}_L - d_0) & \text{if} \quad \mathbf{x}-\mathbf{x}_L \in (d_0; d_1] \\ 0 & \text{if} \quad \mathbf{x}-\mathbf{x}_L > d_1 \end{cases} \tag{9}$$

The resulting maximum likelihood estimates are shown in Table 1 and Table 2. The estimate of ϕ in the piecewise constant model lies on the border of the sample space, $\hat{\phi} = 0$, but a plot of the loglikelihood (not shown) as a function of ϕ keeping all other parameters fixed at the maximum likelihood estimates in Table 1 shows that the maximum is indeed reached at 0. With the continuous model the estimate for ψ also lies on the border of the sample space, and a plot of the loglikelihood contours (not shown) as a function of ψ and ϕ keeping the other parameters fixed at the maximum likelihood estimates shows that the maximum is indeed reached at 0. Figure 6 shows the maximum loglikelihood values for each pair in the (d_0, d_1)-grid with the red square showing the maximum value. The maximum is unique, but as seen in the figure several d_0-estimates gave loglikelihood values

very close to the maximum, indicating that this parameter was poorly estimated. Testing no effect of cumulative call time $\hat{\phi} = 0$ gave the likelihood ratio test statistic $D = 0.000$ for the piecewise constant decreasing model and $D = 1.53$ for the continuous model. The χ^2_1-approximation to the null distribution did not hold in any case, and 1000 simulations (less than 1% did not converge for the continuous model) gave respectively the empirical $p = 0.522$ and $p = 0.401$. As expected from Figure 5 we cannot reject the hypothesis of call time not affecting the distance relation.

It is preferable to include covariates in the continuous model but our data fit this poorly as seen in the estimates for the d-parameters, and in such a case it can be relevant to use the piecewise constant model instead. Adjustment for covariates influencing the spatial baseline distribution can obviously also be added to the distance relation.

5 Discussion

It is relevant to include tumour localisation in the model, when modelling the association between brain tumours and the use of mobile phones, and this might add interesting knowledge about the association. We have shown how a spatial point process model can be used to investigate the association between brain tumours and mobile phones taking the tumour localisation into consideration. Our approach has the advantage of only using cases eliminating the differential bias between cases and controls, and of using the specific tumour localisations instead of the more cruder approaches using brain regions[20] or a dichotomised exposure measure based on the distances calculated from the localisations[11].

The approach requires all included subjects to be either labelled as a left-user or a right-user, which is a limitation since all users did not have a preferred side. Furthermore the information on preferred side is self-reported and exposed to reporting bias, since cases were aware of which side of the head their tumour occured when answering the question on preferred side of phone use, which raises concern about the direction of association[30, 31]. Our suggestion on how to deal with the uncertainty about preferred side is to write the intensities for left- and right-users as functions of both the distance to the right ear and the distance to the left ear. The two distance functions in each intensity should then be assigned weights reflecting which side was preferred. This approach can also be used to include the persons not preferring one side by setting both weights equal to 0.5.

Before applying the point process model to our data we started with nonparametric density estimation for the left-side users and the right-side users seperately as an explorative analysis. We did not use edge correction for this, but it was seen in Figure 1 that the kernels smooth off before reaching the edge of the brain, and therefore edge correction is not crucial here. The distance relation in the point process model described by the function g was modelled in various ways. All our specifications of g only considered radial distance effects from the phone. This is in agreement with how RF energy is distributed through the brain, but with other exposure sources directional preference in the spatial effects[32] might be relevant to consider. The shortest distance between tumour and preferred ear in our data was 39 mm and therefore it was not possible to estimate g close to the ear. This is the main reason why we did not model g as a strictly monotone continuous function, though others have used the exponential function to describe the absorption of RF energy from

mobile phones in the brain[33]. The continuous distance function we modelled resulted in estimates hard to interpret, but this does not imply general problems with the model, but is again a consequence of how our data were spatially distributed. We included the covariate cumulative phone use in the model in (8) and in (9) but of course covariates can be included differently. The choice of how to include a covariate should depend on how it is thought to influence the distance relation. If a high level of cumulative phone use was assumed to have a higher influence on the intensity than a low, then $(1+\phi z_{ph})$ in (8) could be replaced by $(\exp(\phi z_{ph}))$. Another assumption could be that the influence of phone use decreased with the distance and then the expression in (8) would be $(1+\phi \cdot z_{ph}/(\mathbf{x}-\mathbf{x}_L))$. Hence, it is possible to include various covariates in the model. Typical covariates such as age and sex that affect the prevalence of gliomas[34] should be included as an adjustment if they are thought to also influence the spatial baseline distribution of gliomas.

We have shown for our motivating data that distance to the preferred ear for mobile phone use had an influence on the intensity of brain tumours for mobile phone users. However, we have not used the full Interphone data but only the subset of gliomas where a single voxel was marked by neuroradiologists as the tumour origin and the person reported either left- or right-side as preferred for mobile phone use.

Acknowledgements

Funding: This work was fully supported by funding from the Danish Cancer Society's scientific committee (KBVU), grant R20-A897. The Interphone study was supported by

funding from the European Fifth Framework Program, 'Quality of Life and Management of Living Resources' (contract QLK4-CT-1999901563) and the International Union Against Cancer (UICC). The UICC received funds for this purpose from the Mobile Manufacturers' Forum and GSM Association. Provision of funds to the Interphone study investigators via the UICC was governed by agreements that guaranteed Interphone's complete scientific independence. Most of the Interphone centers received some additional national complementary funding for the conduct of Interphone, with a complete list published in: The Interphone Study Group. Brain tumour risk in relation to mobile telephone use: results of the Interphone international case-control study. Int J Epidemiol 2010; 39: 675–94.

Thanks: The authors like to thank Mrs Monika Moissonier (IARC) for extracting the data used for this work from the Interphone database. The authors like to thank the principal investigators of the national Interphone components for the provision of data for this project: Drs Bruce Armstrong, Australia, Jack Siemiatycki, Canada - Montreal, Dan Krewski, Canada - Ottawa, Mary McBride, Canada - Vancouver, Christoffer Johansen, Denmark, Anssi Auvinen, Finland, Martine Hours, France, Joachim Schüz, Germany, Siegal Sadetski, Israel, Susanna Lagorio, Italy, Naohito Yamaguchi, Japan, Alistair Woodward, New Zealand, Tore Tynes, Finland, Maria Feychting, Sweden, Patricia McKinney, UK-North, Anthony J Swerdlow, UK-South. The authors like to thank the IARC team coordinating the study during its fieldwork: Drs Elisabeth Cardis, Isabelle Deltour, Lesley Richardson, and Martine Vrijheid, and Mrs Monika Moissonnier, Mrs Emilie Combalot and Mrs Helen Tardy. Finally, the authors like to thank Johanna Vompras who programmed the Gridmaster programme under the supervision of the German Interphone team and the

IARC coordinators.

References

1. Diggle PJ, Gatrell AC, Lovett AA. Modelling the prevalence of cancer of the larynx in part of Lancashire: A new methodology for spatial epidemiology. *Spatial Epidemiology*, Thomas RW (ed.). Pion: London, 1990; 35–47.

2. Diggle PJ. A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1990; **153**(3):349–362, doi:10.2307/2982977.

3. Cook-Mozaffari PJ, Darby SC, Doll R, Forman D, Hermon C, Pike MC, Vincent T. Geographical variation in mortality from leukaemia and other cancers in England and Wales in relation to proximity to nuclear installations, 1969–78. *British Journal of Cancer* 1989; **59**(3):476–485, doi:10.1038/bjc.1989.99.

4. Bithell JF, Stone RA. On statistical methods for analysing the geographical distribution of cancer cases near nuclear installations. *Journal of Epidemiology and Community Health* 1989; **43**(1):79–85, doi:10.1136/jech.43.1.79.

5. Elliott P, Hills M, Beresford J, Kleinschmidt I, Jolley D, Pattenden S, Rodrigues L, Westlake A, Rose G. Incidence of cancers of the larynx and lung near incinerators of waste solvents and oils in Great Britain. *The Lancet* 1992; **339**(8797):854–858, doi:10.1016/0140-6736(92)90290-J.

6. Michaelis J, Keller B, Haaf G, Kaatsch P. Incidence of childhood malignancies in the vicinity of West German nuclear power plants. *Cancer Causes & Control* 1992; **3**(3):255–263, doi:10.1007/BF00124259.

7. Muirhead C, Darby S. (eds). Royal Statistical Society Meeting on cancer near nuclear installations. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1989; **152**(3):305–384, doi:10.2307/2983128.

8. Diggle PJ, Rowlingson BS. A conditional approach to point process modelling of elevated risk. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1994; **157**(3):433–440, doi:10.2307/2983529.

9. Diggle PJ. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press: Boca Raton, 2013.

10. Hartikka H, Heinävaara S, Mäntylä R, Kähärä V, Kurttio P, Auvinen A. Mobile phone use and location of glioma: A case-case analysis. *Bioelectromagnetics* 2009; **30**(3):176–182, doi:10.1002/bem.20471.

11. Larjavaara S, Schüz J, Swerdlow A, Feychting M, Johansen C, Lagorio S, Tynes T, Klaeboe L, Tonjer SR, Blettner M, *et al.*. Location of gliomas in relation to mobile telephone use: A case-case and case-specular analysis. *American Journal of Epidemiology* 2011; **174**(1):2–11, doi:10.1093/aje/kwr071.

12. Takebayashi T, Varsier N, Kikuchi Y, Wake K, Taki M, Watanabe S, Akiba S, Yamaguchi N. Mobile phone use, exposure to radiofrequency electromagnetic field, and

brain tumour: a case-control study. *British Journal of Cancer* 2008; **98**(3):652–659, doi:10.1038/sj.bjc.6604214.

13. Cardis E, Armstrong BK, Bowman JD, Giles GG, Hours M, Krewski D, McBride M, Parent ME, Sadetzki S, Woodward A, *et al.*. Risk of brain tumours in relation to estimated RF dose from mobile phones: results from five Interphone countries. *Occupational and Environmental Medicine* 2011; **68**(9):631–640, doi:10.1136/oemed-2011-100155.

14. Cardis E, Richardson L, Deltour I, Armstrong B, Feychting M, Johansen C, Kilkenny M, McKinney P, Modan B, Sadetzki S, *et al.*. The INTERPHONE study: design, epidemiological methods, and description of the study population. *European Journal of Epidemiology* 2007; **22**(9):647–664, doi:10.1007/s10654-007-9152-z.

15. Larjavaara S, Mäntylä R, Salminen T, Haapasalo H, Raitanen J, Jääskeläinen J, Auvinen A. Incidence of gliomas by anatomic location. *Neuro-Oncology* 2007; **9**(3):319–325.

16. Inskip PD, Tarone RE, Hatch EE, Wilcosky TC, Selker RG, Fine HA, Black PM, Loeffler JS, Shapiro WR, Linet MS. Laterality of brain tumors. *Neuroepidemiology* 2003; **22**(2):130–138.

17. Vrijheid M, Deltour I, Krewski D, Sanchez M, Cardis E. The effects of recall errors and of selection bias in epidemiologic studies of mobile phone use and cancer risk. *Journal of Exposure Science and Environmental Epidemiology* 2006; **16**(4):371–384, doi:10.1038/sj.jes.7500509.

18. Vrijheid M, Armstrong BK, Bédard D, Brown J, Deltour I, Iavarone I, Krewski D, Lagorio S, Moore S, Richardson L, *et al.*. Recall bias in the assessment of exposure to mobile phones. *Journal of Exposure Science and Environmental Epidemiology* 2009; **19**(4):369–381, doi:10.1038/jes.2008.27.

19. Düsseldorf: Vompras. *GridMaster Computer Program* 2007.

20. The INTERPHONE Study Group. Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. *International Journal of Epidemiology* 2010; **39**(3):675–694, doi:10.1093/ije/dyq079.

21. Bit-Babik G, Chou CK, Faraone A, Gessner A, Kanda M, Balzano Q. Estimation of the SAR in the human head and body due to radiofrequency radiation exposure from handheld mobile phones with hands-free accessories. *Radiation Research* 2003; **159**(4):550–557, doi:10.1667/0033-7587(2003)159[0550:EOTSIT]2.0.CO;2.

22. Cardis E, Deltour I, Mann S, Moissonnier M, Taki M, Varsier N, Wake K, Wiart J. Distribution of RF energy emitted by mobile phones in anatomical structures of the brain. *Physics in Medicine and Biology* 2008; **53**(11):2771–2783, doi:10.1088/0031-9155/53/11/001.

23. Duong T. *ks: Kernel smoothing* 2013. URL `http://CRAN.R-project.org/package=ks`, R package version 1.8.13.

24. Silverman BW. *Density estimation for statistics and data analysis*. CRC Press: Boca Raton, 1986.

25. Duong T. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *Journal of Statistical Software* 10 2007; **21**(7):1–16.

26. Deltour I, Wiart J, Taki M, Wake K, Varsier N, Mann S, Schüz J, Cardis E. Analysis of three-dimensional SAR distributions emitted by mobile phones in an epidemiological perspective. *Bioelectromagnetics* 2011; **32**(8):634–643, doi:10.1002/bem.20684.

27. Rodrigues A, Diggle P, Assuncao R. Semiparametric approach to point source modelling in epidemiology and criminology. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 2010; **59**(3):533–542, doi:10.1111/j.1467-9876.2009.00708.x.

28. Wood S. *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation* 2014. URL `http://CRAN.R-project.org/package=mgcv`, R package version 1.7.29.

29. Dimbylow PJ, Mann SM. SAR calculations in an anatomically realistic model of the head for mobile communication transceivers at 900 MHz and 1.8 GHz. *Physics in Medicine and Biology* 1994; **39**(10):1537–1553, doi:10.1088/0031-9155/39/10/003.

30. Schüz J. Lost in laterality: interpreting "preferred side of the head during mobile phone use and risk of brain tumour" associations. *Scandinavian Journal of Public Health* 2009; **37**(6):664–667, doi:10.1177/1403494809341096.

31. Frederiksen K, Deltour I, Schüz J. Estimating associations of mobile phone use and brain tumours taking into account laterality: a comparison and theoretical evaluation

of applied methods. *Statistics in Medicine* 2012; **31**(28):3681–3692, doi:10.1002/sim.5425.

32. Lawson AB. On the analysis of mortality events associated with a prespecified fixed point. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1993; **156**(3):363–377, doi:10.2307/2983063.

33. Wake K, Varsier N, Watanabe S, Taki M, Wiart J, Mann S, Deltour I, Cardis E. The estimation of 3D SAR distributions in the human head from mobile phone compliance testing data for epidemiological studies. *Physics in Medicine and Biology* 2009; **54**(19):5695–5706, doi:10.1088/0031-9155/54/19/003.

34. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK. *WHO classification of tumours of the central nervous system*. IARC: Lyon, 2007.

Tables

Estimate (se)	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\phi}$
g piecewise const.	1.33 (0.46)	2.03 (0.41)	1.50 (0.26)	1.13 (0.25)	-
g piecewise const., decr.	1.82 (0.33)	1.82 (0.29)	1.48 (0.23)	1.09 (0.17)	-
g piecewise const., decr. with phone use	1.82 (0.33)	1.82 (0.31)	1.48 (0.24)	1.09 (0.12)	0.00 (0.33)

Table 1: Maximum likelihood estimates and Monte Carlo standard errors with piecewise constant g functions

Estimate (se)	$\hat{\psi}$	$\hat{d}_0$	$\hat{d}_1$	$\hat{\phi}$
g continuous	0.656 (0.78)	90 (-)	100 (-)	-
g continuous with phone use	0.00 (0.64)	45 (-)	100 (-)	1.78 (1.03)

Table 2: Maximum likelihood estimates and Monte Carlo standard errors with continuous g function

Figures

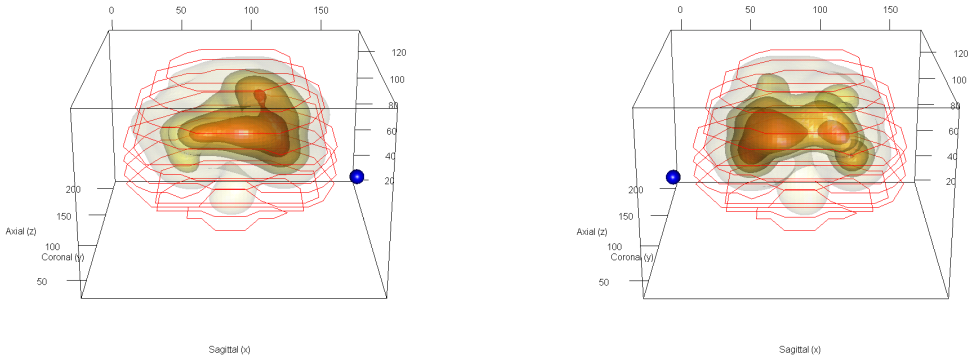

Figure 1: *Left*: Probability density for the left-side phone users shown as nested 3-d contours at 25%, 50%, 75% and 100% as the upper percentages of highest density regions. The red polygons bound the brain, the blue point is the left ear. *Right*: Similar figure for the right-side users with right ear shown as blue point.

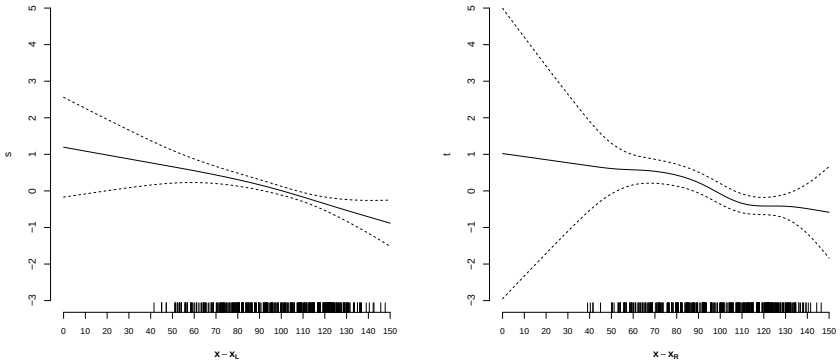

Figure 2: *Left*: The smooth function of the distances to the left ear. The rug marks on the x-axis show the distance to left ear for both left- and right-users. *Right*: Similar figure for the distances to right ear.

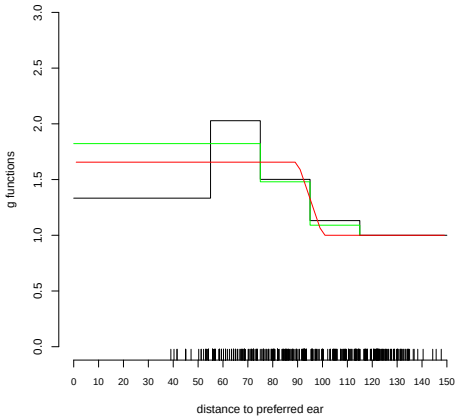

Figure 3: Estimated g functions: piecewise constant, piecewise constant and declining, and continuous. The rug marks on the x-axis show the distances to preferred ear.

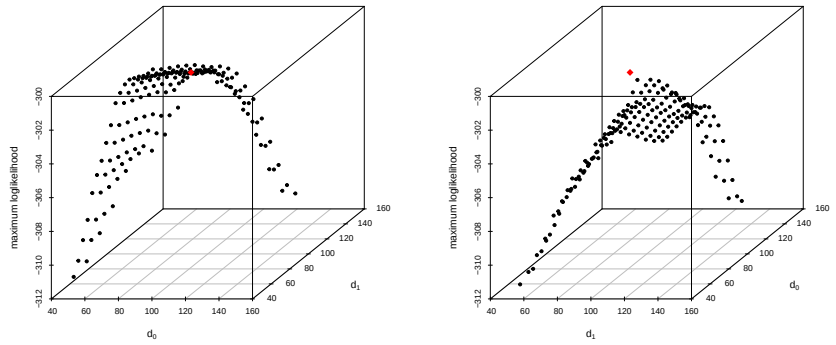

Figure 4: *Left*: Maximum loglikelihood values for grid of (d_0, d_1)-values. The maximum value is marked as red. *Right*: Exchanged d_0 and d_1 on the axis for different view.

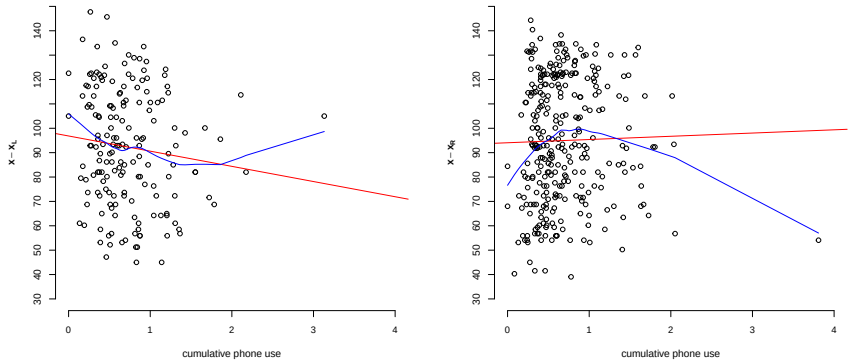

Figure 5: *Left*: Distance to left ear plotted against transformed phone use for the left-users. Red line is best linear fit and blue curve a smoother using locally-weighted polynomial regression. *Right*: Similar for right-users.

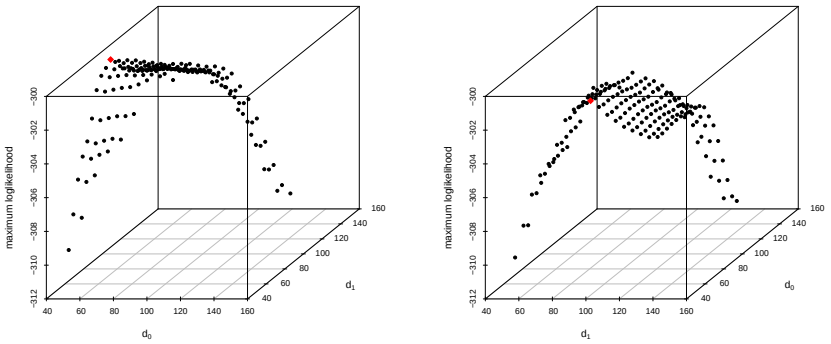

Figure 6: *Left*: Maximum loglikelihood values for grid of (d_0, d_1)-values for model including phone use. The maximum value is marked as red. *Right*: Exchanged d_0 and d_1 on the axis for different view.