

Original Paper

GM Kendall, P Chernyavskiy, JD Appleton, JCH Miles, R Wakeford, M Athanson, TJ Vincent, NP McColl, MP Little

Modelling the bimodal distribution of indoor gamma-ray dose-rates in Great Britain

GM Kendall (✉)

Cancer Epidemiology Unit, University of Oxford, Richard Doll Building, Old Road Campus, Headington, Oxford, OX3 7LF, UK.

Email: Gerald.Kendall@ceu.ox.ac.uk or Gerald.Kendall@cantab.net

Tel: +44 (0)1865 289600

Fax: +44 (0)1865 289610

P Chernyavskiy

Radiation Epidemiology Branch, National Cancer Institute, DHHS, NIH, Division of Cancer Epidemiology and Genetics, Bethesda, MD 20892-9778, USA

Present address: Department of Mathematics and Statistics, University of Wyoming, Laramie, WY 82071, USA

J D Appleton

British Geological Survey, Kingsley Dunham Centre, Nicker Hill, Keyworth, Nottingham, NG12 5GG, UK

JCH Miles

49 Nobles Close, Grove, Oxfordshire, OX12 0NR UK

R Wakeford

Centre for Occupational and Environmental Health, Institute of Population Health, The University of Manchester, Ellen Wilkinson Building, Oxford Road, Manchester, M13 9PL, UK

M Athanson

Bodleian Library, University of Oxford, Broad Street, Oxford, OX1 3BG, UK

TJ Vincent

Childhood Cancer Research Group, University of Oxford, New Richards Building, Old Road, Oxford, UK

NP McColl

Centre for Radiation, Chemical and Environmental Hazards, Public Health England, Chilton, Didcot Oxon, OX11 0RQ, UK

MP Little

Radiation Epidemiology Branch, National Cancer Institute, DHHS, NIH, Division of Cancer Epidemiology and Genetics, Bethesda, Maryland 20892-9778, USA

Abstract Gamma radiation from naturally occurring sources (including directly ionising cosmic-rays) is a major component of background radiation. An understanding of the magnitude and variation of doses from these sources is important and the ability to predict them is required for epidemiological studies. In the present paper, indoor measurements of naturally occurring gamma-rays at representative locations in Great Britain are summarized. It is shown that although the individual measurement data appear unimodal, the distribution of gamma-ray dose-rates when averaged over relatively small areas, which probably better represents the underlying distribution with inter-house variation reduced, appears bimodal. The dose-rate distributions predicted by three empirical and geostatistical models are also bimodal and compatible with the distributions of the areally-averaged dose-rates. The distribution of indoor gamma-ray dose-rates in the UK is compared with those in other countries, which also tend to appear bimodal (or possibly multimodal). The variation of indoor gamma-ray dose-rates with geology, socio-economic status of the area, building type and period of construction are explored. The factors affecting indoor dose-rates from background gamma radiation are complex and frequently intertwined, but geology, period of construction and socio-economic status are influential; the first is potentially most influential, perhaps because it can be used as a general proxy for local building materials. Various statistical models are tested for predicting indoor gamma-ray dose-rates at unmeasured locations. Significant improvements over previous modelling are reported. The dose-rate estimates generated by these models reflect the imputed underlying distribution of dose-rates and provide acceptable predictions at geographical locations without measurements.

Introduction

Exposures to terrestrial gamma-rays and to cosmic-rays are important components of the typical individual radiation doses received in every country. Most people spend much of their time in buildings and gamma-ray exposures indoors are often larger than those outdoors. Nevertheless, outdoor measurement data are usually more available than those for indoor exposures (UNSCEAR, 2010). We recently published an analysis of naturally occurring gamma radiation in British homes (Kendall *et al*, 2016b). This analysis was of population-based sample measurements and included both terrestrial gamma-rays from radionuclides in rocks, soils and building materials, and also the directly ionizing component of cosmic-rays, but not cosmic-ray neutrons. The present paper discusses the same types of radiation, which will be referred to as “gamma-rays” for brevity, and differs from its predecessor in a much more wide-ranging investigation of methods for predicting and interpreting gamma-ray dose-rates in unmeasured buildings, and in the conclusions drawn about the nature of their distribution.

The indoor gamma-ray dose-rate in a particular dwelling may be regarded as the dose-rate in the open air before the building was constructed, reduced by the shielding provided by the material of the house, but then increased by radiation from the building materials. As described by Kendall and co-workers. (Kendall *et al*, 2016b), dose-rates in neighbouring houses tend to be similar, but there is considerable inter-house variation. There are complex reasons for the similarity of dose-rates in neighbouring houses. The most important are probably that nearby houses will tend to be built of similar materials and to similar designs, and the role played by the local geology, particularly if building materials are locally sourced.

The previous paper (Kendall *et al*, 2016b) described in detail the provenance of the gamma-ray dose-rate measurements and the pattern of accumulated doses. It also considered methods that might be used to estimate dose-rates in unmeasured locations. An investigation of such methods applied to dose-rates collected in British homes found that the best results came from an *ad hoc* weighted linear combination of nearest measurements and similar simple estimates. Interpolation of indoor gamma dose-rates has also been performed using kriging models, i.e., predictive models employing a formal spatial structure. Warnery and co-workers (Warnery *et al*, 2015) compared two variogram-based kriging models estimated using dose-rates in France. Chernyavskiy and co-workers. (Chernyavskiy *et al*, 2016) examined several variogram- and likelihood-based modelling approaches with respect to

interpolating dose-rates in Great Britain (GB: England, Wales and Scotland). One such method used a new class of multi-resolution Gaussian process (MRGP) models (Nychka *et al*, 2015), which approximate the true (but unknown) Gaussian process with a varying number of stochastic sums of Wendland compactly supported spatially-symmetric radial basis functions (Wendland, 1995). At each level the scaling of the component basis functions is reduced by a factor of 2, so that progressively larger numbers of such basis functions are required at each level (increasing roughly by a factor of 2 at each level) to cover the spatial grid (Nychka *et al*, 2015). Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) noted that because MRGP models are fitted using maximum likelihood, there is a more robust framework for likelihood-based inference which is expected to provide more reliable estimates of uncertainty for the predicted dose-rates; they concluded that these models could be considered for use in epidemiological studies, in addition to the linear combination method identified by Kendall and co-workers (2016b). More details are given in Materials and Methods and in the publications cited (Kendall *et al*, 2016b) (Chernyavskiy *et al*, 2016).

Kendall and co-workers (Kendall *et al*, 2015) (Kendall *et al*, 2016a) reported a substantial variation of indoor radon concentrations with socio-economic status (SES) in GB, with residential radon levels experienced by people living in the most deprived areas being, on average, only about two-thirds of those of the population as a whole. This correlation was seen using both areal and individual measures of SES, and both for directly measured and for modelled radon concentrations. Kendall and co-workers (Kendall *et al*, 2015) reported a much smaller variation of indoor gamma-ray dose-rates with areal SES, the mean residential gamma-ray dose-rate rising from about 91 nGy/h in Carstairs Quintile 1 (least deprived) areas to about 99 nGy/h in Quintile 5. Methods for estimating indoor gamma-ray dose-rates on a residence-specific basis were not available to Kendall and co-workers (2015), and means for intermediately-sized areal units, 459 county districts, were used (based on 10,199 measurements). Although SES therefore appears to play a relatively minor role in determining indoor gamma-ray dose-rates this topic will be explored in more detail.

In this paper the distribution of indoor gamma-ray dose-rates in the UK is assessed, using simple areal averaging to extract any patterns in the data, and explore the variation with geology, SES and period of construction in order to throw light on the underlying distribution. The “underlying” distribution of indoor gamma-ray dose-rates is an idealised distribution in which differences in construction layout and materials between neighbouring buildings are averaged out while allowing regional variations to persist. This is not of particular practical importance but is of scientific interest since it reflects the

fundamental factors that determine indoor gamma-ray dose-rates. It is likely that model fits to observed data will reflect this underlying distribution. Then, the observed indoor gamma-ray dose-rates are compared with the predictions of the models referred to above. The reported distributions of indoor gamma-ray dose-rates in other countries are also reviewed. The paper thus has two aims: 1) to describe the distribution of indoor gamma-ray dose-rates in Great Britain and to explore the factors that influence it; and 2) to develop and test statistical models to predict indoor gamma-ray dose-rates at unmeasured locations.

Materials and methods

Indoor gamma-ray measurement data

Residential indoor gamma-ray dose-rate measurement data are from two sources:

- a) The UK National Survey of naturally occurring radiation exposures in dwellings (Wrixon *et al*, 1988), including measurements in 2,283 dwellings; and
- b) The UK Childhood Cancer Study, UKCCS (UK Childhood Cancer Study Investigators, 2000) (UK Childhood Cancer Study Investigators, 2002), including measurements in 7,916 dwellings.

The total set of 10,199 indoor gamma radiation measurements from these two sources was used as the basis for the analyses of Kendall and co-workers (Kendall *et al*, 2016b) and Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016). In the UK National Survey (Wrixon *et al*, 1988), the sample of dwellings chosen was population weighted. The UKCCS was based on an essentially complete sample of homes in which a child developed cancer over the period of study together with those of matched population-based controls, and these residences too should reasonably closely follow the distribution of the general population. However, it should be noted that both for the National Survey and for the UKCCS, measurements could be completed at only about half the addresses originally selected and some participation bias is to be expected. For both series the measurements were made using long-term thermoluminescent detectors (TLDs). In over 97% of instances the dose-rate assigned to a house was the weighted mean of the results from two dosimeters placed in different parts of the dwelling. Further details are given by Kendall and co-workers (Kendall *et al*, 2016b). The uncertainty quoted for any dosimeter reading was largely due to uncertainties in the angular and energy variation of the

incident radiation (Wrixon *et al*, 1988) (Shaw & Wall, 1977) (Kendall *et al*, 2016b). The uncertainty in measurements of environmental gamma-rays was probably around 5% (Driscoll *et al*, 1983) (Kendall *et al*, 2016b).

Outdoor Gamma-ray dose rates

Information on external (outdoor) gamma-ray dose rates came from a survey by Green and co-workers (Green *et al*, 1989): 2,398 measurements were made using an energy-compensated Geiger-Müller tube over a period of ten minutes. The principal published results were the double-smoothed dose-rates tabulated for each 10 km-side square of the National Grid in GB. However, estimates were sparse in some coastal areas and in islands, and Kendall and co-workers (Kendall *et al*, 2013) interpolated values for 4% of the squares, containing 0.5% of the subjects included in the record-based case-control study.

The population-weighted average outdoor dose-rate in GB is about 77 nGy/h, somewhat lower than the indoor value. This is because in a majority of cases within a dwelling the shielding of external radiation is more than offset by radiation from the building materials. These outdoor dose-rates were correlated with the measured indoor gamma-ray dose rates, although this correlation is not particularly strong (Pearson correlation coefficient=0.312, 95% Confidence Interval, CI 0.294 - 0.329). However, it is possible that the correlation would have been stronger had there been available individual outdoor measurements at effectively the same locations as the indoor measurements.

Geological data

Geological classifications, together with SES and other census-based data, were used to improve the models for predicting indoor gamma-ray dose-rates. These were broadly as used previously (Chernyavskiy *et al*, 2016; Kendall *et al*, 2016b). In the present study use was made of the sixteen geological “Bedrock Classes” used by Kendall and co-workers (Kendall *et al*. 2016b). Chernyavskiy and co-workers (Chernyavskiy *et al*. 2016) introduced a further degree of simplification with a set of six “Reduced Bedrock Classes”. These consist of five combinations of geological periods and granite. The data and the small changes since the previous analyses are summarized in Online Resource 1. Parameters of the geological classifications are summarized in Table 1.

Census data

In examining the influence of SES on indoor gamma-ray dose-rates use was made of a number of parameters from the UK national censuses and also the Carstairs index of deprivation (Carstairs and Morris 1991). This is assessed for census wards (small areas with a mean all-ages population of about 5,000). More details of the Carstairs scores and other census-derived parameters such as urban/rural status are to be found in Online Resource 1.

Period of construction of dwellings

Building practices change with time and it will be shown that the period in which a dwelling was constructed is useful in predicting the gamma-ray dose-rate within it. In some countries there is a register of buildings which records this information, but unfortunately this does not apply to GB. However, an approximate division into buildings probably constructed before or after 1940 is possible (Appleton & Cave, 2018). In the 1930s detailed maps of the land use in GB were prepared under the direction of L Dudley Stamp (Dudley Stamp, 1931). The sheets relating to England and Wales have been prepared for use in a geographical information system (GIS) (Environment Agency, 2010) (Southall *et al.*, 2007), though most of those for Scotland have not. The original 14-level Dudley Stamp land use classifications were simplified to eight levels in the GIS versions; two of these were “Urban” and “Suburban”. For purposes of the Pre-/Post-1940 construction classification it is assumed that dwellings in areas classified as urban or suburban on the Dudley Stamp maps were in existence in 1940 while those on land used for other purposes were built later. This will not accommodate, for example, redevelopment and infill, but the classification will be shown to have predictive value. The mean indoor gamma-ray dose-rates in dwellings classified as Pre- and Post-1940 were 97.1 and 94.2 nGy/h, respectively. More details, including the extension to Scotland, are given in Appendix A.

Birth locations of participants in an epidemiological study

Here the distributions of doses calculated in various ways at the birth locations of study subjects in a large epidemiological study are discussed. These birth locations follow the distribution of births in the general population, but the birth sample is an order of magnitude larger than the set of measurement locations. The study subjects were from a large matched case-control study of childhood cancer and natural background radiation (Kendall *et al.* 2013), but with an expanded range of calendar years (1962 to 2010 rather than 1980 to 2006). The enhanced dataset included 54,462 cases and 69,992 controls, giving a total of 124,454 study subjects. This is slightly smaller than the population

considered by Kendall and co-workers (Kendall *et al.* 2016b) (126,817) because records from incomplete case-control sets are now excluded.

Statistical Methods

Kendall and co-workers (Kendall *et al.*, 2016b) explored a number of simple methods for estimating dose-rates in unmeasured residences. These included

- averages over relatively small areas (for example, 10 km squares, county districts or postcode districts),
- averages over geological units of various kinds,
- sums of the readings at the nearest measurement points, weighted by distance in various ways (“weighted sum of nearest neighbours”)

All these methods made reasonably good predictions of dose-rates and all were significantly better than simply assigning the national average to unmeasured locations. The weighted sum of nearest neighbour measurements was the best of the simple methods investigated, although not strikingly so. Grouping by geological unit was not notably more effective than grouping by other areas and Kendall and co-workers (Kendall *et al.*, 2016b) concluded that complex factors resulted in indoor gamma-ray dose-rates tending to be similar in neighbouring buildings. The best estimates of dose-rates in unmeasured dwellings were obtained as a weighted linear combination of nearby indoor measurements using ordinary least squares (the “extended OLS (E-OLS) model”). The selection of the optimum model from the possible set of candidate models was made using the Akaike Information Criterion (Akaike, 1973; Akaike, 1981) as described in Appendix B.

A spatial geostatistical model was also fitted to the gamma-ray dose-rates, which assumed an underlying geologically-determined spatial variation, superimposed on which is a stochastic process that models the observed tendency of dose-rates in neighbouring houses to correlate, more than can be accounted for by purely geological factors (Diggle & Ribeiro, 2007). Specifically, it was assumed that the gamma-ray dose-rate at a given location was given by a sum of: (a) a “mean” process, with distinct values given by a combination of various geological and spatial variables; (b) a stochastic spatial correlation process; and (c) a Gaussian “noise” term. The spatial correlation was assumed to result from unobserved variables that would be correlated between nearby areas (with greater correlation the closer the areas are), and that might affect gamma-ray dose-rate. The spatial correlation process was

that developed by Matérn (Matérn, 1960); further details are given in Appendix C. In (Kendall *et al*, 2016b) the only such model considered used a 16-level 50K-BEDSUP-bedrock variable to describe the mean process; in the present paper this variable is considered but also various other geological and spatial variables, as outlined in Appendix C. The Gaussian “noise” term was assumed to be the result of random measurement error in assessing the dose-rates. The parameters of the model were estimated by maximum likelihood in R (R Project version 3.4.4, 2018). The optimal model was chosen, as indicated in Appendix C, as that which approximately minimised the Akaike Information Criterion (Akaike, 1973), guided in part by a series of likelihood ratio tests; however, an exhaustive investigation of the available submodels was not performed. As well as the location variables, Easting and Northing, the optimal Gaussian- Matérn model includes the Dudley Stamp Pre-/Post-1940 construction category, the urban-rural (6-level) classification, the external gamma-ray dose-rate and the 50K-BEDSUP Surface geological classification (23-level).

The present paper also makes use of geospatial models developed by Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) who examined a number of variogram-based and MRGP models. An MRGP model with eight components (MRGP-8) and a variogram-based Spherical model performed best, although neither fitted the data quite as closely as the E-OLS model described above. However, the MRGP model has the advantage of allowing more reliable estimates of uncertainty than the E-OLS model. Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) proposed that the MRGP-8 and Spherical models could be considered alongside the E-OLS model for use in epidemiological studies. However, intractable numerical problems arose in applying the Spherical model to a dataset as large as that of the birth locations and, therefore, attention here is focussed on the MRGP-8 models.

In the present paper, two sets of spatially-varying covariates are used in the 8-level MRGP models. The 3-covariate model (MRGP-8-3) has covariates consisting of: Latitude, Longitude, and External Gamma-ray Dose-rates, and the 5-covariate MRGP model (MRGP-8-5) has covariates consisting of: Latitude, Longitude, External Gamma-ray Dose-rates, Carstairs Score in 1981, and Population Density in 1981. (This differs from the covariates used in the MRGP model of Chernyavskiy *et al* (2016), which were Carstairs Score in 1981, Population Density in 1981 and Urban/Rural Status in 1981, although this last categorical variable could not be used in the modelling.) The inclusion of categorical variables in MRGP models was not possible because the design matrix would become singular; thus, only the continuous variables were included. The models were implemented using the LatticeKrig (Nychka, 2014) package in R (R Core Team, 2017).

More details of the statistical modelling are to be found in Appendices B and C. Note, in particular, that while the modelling of (Kendall *et al*, 2016b) took into account the location of measurements, administrative areas such as county districts and the geology of the location, the calculations presented here also consider a variety of census-based SES parameters, period of construction of the building and the outdoor gamma-ray dose-rate at the location in question.

In this paper distributions of grouped dose-rates are reported by area. This means that all the measurements in a small area were regarded as the mean for that area. While this involves just grouping and averaging, it may also be described as the dose-rates predicted by fitting a linear model of dose-rate using indicators for the areas as predictors, i.e., in which the dose-rate, DR_i , for individual i , associated with area group, $area_i$, is given by

$$DR_i = \sum_{j=1}^N \beta_j 1_{area=j} + \varepsilon_i \quad (1)$$

for some parameters $(\beta_j)_{j=1}^N$ and independent identically distributed Normal errors, ε_i .

In the present paper, the dose-rate distribution is also described more formally as a superposition of Normal distributions, as described in Appendix D.

The performance of the models was evaluated using various statistics. As well as the predicted overall mean and the mean absolute error (i.e., the mean of the absolute difference between measurements and predictions) how effectively the models predicted the actually observed gamma dose rates is considered using the Pearson (Pearson, 1895) and Spearman (Spearman, 1904) correlation coefficients. The model-fitted standard deviation (i.e., the sample standard deviation of the model-fitted dose-rates) is also evaluated, as well as the predictive Mean Square Error (MSE), given by the average of $[\text{true dose-rate} - \text{model predicted dose-rate}]^2$ over the 30% test sample of the respective models fitted to the randomly chosen 70% model-fit sample; the same 30:70 samples were used for all models.

Results

Both for the National Survey and for the UKCCS the individual measurements were distributed unimodally and approximately symmetrically with a peak at ~95 nGy/h. The distributions are shown in Fig 1. Quartiles of the overall distribution are at 80.2, 95.3 and 110.2 nGy/h. Directly ionizing cosmic-rays were responsible for about one third of the total (Kendall *et al*, 2016b). However, if the raw measurement data are replaced by dose-rate averages for relatively small areas, whether over the 458 county districts (CDs) in GB with at least one measurement, or over the 1,393 10 km grid squares that have measurements, the distributions appear bimodal, with peaks at ~85 nGy/h and ~105 nGy/h, as shown in Fig 2. These data have been smoothed using a running three-point average.

Figure 3 shows the geographical location of measurements in the lowest and highest quartiles of the distribution (i.e., below 80 and above 110 nGy/h); these points lie predominantly in the lower dose-rate part of the lower peak of the areally-averaged distribution and in the higher dose-rate part of the upper peak, respectively. The intermediate dose-rate range contains contributions from both peaks. Broad geographical differences can be seen with, for example, the London area (broadly defined) having low dose-rate levels.

Figure 4 compares the CD-averaged distribution of indoor gamma-ray dose-rates with that predicted by the E-OLS, MRGP-8-3 and MRGP-8-5 models at that location, where the model predictions are at each measurement location. For the E-OLS model the measurement made at that point is ignored; for the MRGP-8 models all points are included

Further details are given in the Online Resources. Online Resource 1 (OLR-1) summarises the geological, construction period and SES information. OLR-2 gives maps showing the locations at which measurements were made together with geological or SES information. OLR-3 shows in tabular form the numbers of locations, mean Carstairs scores and variation of mean indoor gamma-ray dose-rates by Carstairs quintile (i.e., the variation with SES).

Table 2 gives statistics summarising the results of the fitting of the four main models, E-OLS, Gaussian-Matérn, MRGP-8-3 and MRGP-8-5 to the measurements. More details of the E-OLS model are found in Appendix B and more details of the Gaussian-Matérn model in Appendix C, while the MRGP models are described by Chernyavskiy et al (2016). As can be seen, the predictive performance of the E-OLS model, as measured by the MSE using a 70:30 test:retest ratio, is somewhat superior to the other three models, which all have roughly comparable MSE. For comparison, Table 2 also gives results for two simple models assigning a CD average to the location in question: one is the CD

average calculated using all 10,198 measurements of the full measurement dataset excluding the point in question, and the second relates to CD averages calculated in the manner outlined in the previous analysis (Kendall *et al*, 2013) using only the 2,283 measurements of the National Survey at the 7,916 locations at which non-National Survey measurements had been made. The latter estimates the uncertainties in the CD average dose-rates used in the previous epidemiological analysis (Kendall *et al*, 2013). A notable feature is the rather lower predicted SD for the Gaussian-Matérn model (11.55) and that using the CD mean (12.21).

Table 3 presents an Analysis of Variance (ANOVA) for the preferred E-OLS model. The R^2 is 0.3407, therefore suggesting that the model accounted for 34% of the total variation in the dose-rate data. The largest contribution (almost three-quarters of the 34%) comes from the inverse-distance-weighted sum of nearest neighbouring measurements. The next largest contribution comes from the 50K-BEDSUP and the Dudley Stamp Pre-/Post-1940 categorisation (24% and 1.4% of the 34%, respectively).

Since these inter-house variations are essentially random, given the explanatory variables available and the tendency for dose-rates in nearby houses to be similar, it is of interest to explore the underlying variation in dose-rates by averaging the individual measurements across small areas. As shown in Appendix D, the dose-rate distribution is largely given by a combination of three Normal distributions $X_i \sim N(\mu_i, \sigma^2)$ with means at 80.5, 97.5 and 117.5 nGy/h, with weights $p_i = 0.198, 0.706, 0.052$, respectively. The common estimated SD was $\sigma = 19.79$ nGy/h.

The uncertainties in the E-OLS estimates of indoor dose-rates are described in Appendix B. A bootstrap methodology was used because of strong indications of non-Normality in the residuals (see Appendix B Figure B1). Unfortunately, attempts to estimate uncertainties on the MRGP-8 dose-rate predictions were unsuccessful, numerical problems arising even for small subsets of the data.

Discussion

We first consider the distribution of indoor gamma-ray dose- rates in Great Britain and the major factors that influence it. We then describe the models that we have developed to predict indoor gamma-ray dose- rates at unmeasured locations.

Indoor gamma-ray dose- rates in Great Britain and the factors that influence it

General observations on the distribution of indoor gamma-ray dose-rates

Clearly, inter-house variation is a genuine contributor to the distribution of indoor gamma-ray dose-rates. The stripping away of such variation and of measurement error could lead to a rather academic set of component distributions (see Appendix D). However, the areally-averaged distributions of indoor dose-rates, as shown in Fig 2, lead to the conclusion that underlying bimodality (or possibly multimodality) in the population distribution of indoor gamma-ray dose-rates in GB is obscured by inter-house variation and to a lesser extent measurement errors in the original data. The averaging reduces the effect of these uncertainties and allows the underlying distribution to be apparent. As expected, a further consequence of this averaging process is that the resulting distributions are somewhat less dispersed than that of the original measurements, as a comparison of Fig 2 with Fig 1 demonstrates.

In Appendix E the distributions of indoor gamma-ray dose-rates in other countries is discussed. In the majority of the countries investigated the distribution appears bimodal (or possibly more complex), supporting the findings for GB reported here.

As discussed in the Materials and Methods section, the areally-averaged dose-rates may be considered as the model-predicted estimates of dose-rates resulting from the fit of a linear model of dose-rate using area as the explanatory variable. An obvious question is what separates the dose-rates belonging to the lower peak from those belonging to the upper.

The influence of geology on indoor gamma-ray dose-rates

The geological maps of GB, OLR-2b and OLR-2c, show the locations of measurements below 80 nGy/h and above 110 nGy/h, i.e., of measurements that predominantly fall in the lower and upper peaks, respectively. This suggests that low dose-rates are broadly located on the Tertiary, Cretaceous Clay, Jurassic Oxford Clay, Cretaceous and Jurassic geologies. Table 1 shows that, of the 16 Bedrock Classes, these are the five with the lowest mean gamma-ray dose-rates, all being <90 nGy/h; 41% of the measurements are on these geologies. The next four Bedrock Classes in order of increasing dose-rate (Cambrian, Ordovician, Silurian and Precambrian) are rare, accounting for <3% of the measurements, while the following two (Devonian and Jurassic Ironstone; 5% of the measurements) have dose-rates of 96 nGy/h. The subsequent four geologies, Triassic Mercia Mudstones, Carboniferous, Permian-Triassic, Carboniferous Coal, all with mean dose-rates 101-104 nGy/h,

account for just over half of the measurements (51%). Granite has the highest mean dose-rate (117 nGy/h), but it accounts for only 0.5% of the measurements. The geological maps OLR-2d and OLR-2e show the distribution of measurements below 80 nGy/h and above 110 nGy/h in relation to the six Reduced Bedrock Classes; Table 1 shows that higher dose-rates tend to be found in Carboniferous and Permian-Triassic classes while lower dose-rates tend to occur in the Jurassic/Cretaceous/Tertiary class. However, this geological separation is rough and ready, and it is clear that many high dose-rate measurements are found on “low dose-rate” Bedrock geologies and *vice versa*.

It is noted above that building materials are probably one of the major factors in determining indoor gamma-ray dose-rates. Appendix F gives consideration to the extent to which geological factors might be affecting levels of radioactivity in building materials and thus dose-rates.

The influence of socio-economic status on indoor gamma-ray dose rates

Table OLR-3 shows that the modest variation of indoor gamma-ray dose-rates with the SES of small areas reported by Kendall and co-workers (Kendall *et al*, 2015) is reproduced using the measurements themselves and also by the dose-rate estimation procedures described above. There is about a 10% difference in mean dose-rate between the highest and lowest Carstairs quintiles, which is much smaller than the corresponding variation for radon (Kendall *et al*, 2016a).

Table OLR-3 also shows that a greater proportion of measurements is located in census wards of lesser deprivation (i.e. lower Carstairs score) in comparison to the birth locations of study subjects. This is almost certainly a consequence of a higher relative participation in the measurement programme by the more affluent (Kendall *et al*, 2016a), whereas birth records are not so affected, and it will be noted that the difference in the proportions of measurements and births is particularly marked in the most deprived quintile. However, the Carstairs scores for each quintile are similar between the measurement locations and the birth locations so allowance for Carstairs quintile should compensate for this difference in epidemiological analyses.

Maps OLR-2f and OLR-2g show a tendency for higher indoor dose-rate measurements to be recorded in more socio-economically disadvantaged areas and *vice versa*. However, the role of SES as a determinant of indoor gamma-ray dose-rates is closely entwined with that of geology since many of the more deprived population centres lie on Reduced Bedrock Classes Carboniferous and Permian-

Triassic. The reasons for this are complex and beyond the scope of this paper, but may reflect the old industrial areas found in the coalfields and adjacent Carboniferous and Permo-Triassic domains.

Kendall and co-workers (Kendall *et al*, 2016b) concluded that many factors contribute to indoor gamma-ray dose-rate levels. The difference between mean indoor gamma-ray dose-rates in the lowest and highest Carstairs quintiles is less than 15 nGy/h while the difference between the mean for the group of five geologies with the lowest dose-rates and that for the group of five with the highest (41% and 52% of the totals, respectively, see above) is 18 nGy/h. Geology thus appears marginally better in accounting for differences in indoor gamma-ray dose-rate and may therefore be a somewhat more helpful explanatory factor than SES. This is broadly consistent with findings of Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) who show that using the Carstairs score does not improve the predictive performance of spatial models.

Other factors affecting indoor gamma-ray dose-rates

The type of dwelling and the period of construction both affect indoor radiation levels. Some data on these topics were collected in the National Survey by Wrixon and co-workers (Wrixon *et al*, 1988) and by Kendall and co-workers (Kendall *et al*, 2016a). For completeness a brief summary of the National Survey data is given in Appendix A. However, registers of these housing characteristics are not kept in GB and they are thus of limited predictive use. The exception is the rough and ready separation of dwellings into pre- and post-1940 construction, also described in Appendix A. This, based on land-use maps from the 1930s, has useful predictive power for indoor gamma-ray dose-rates (see Appendices B and C).

Observed and modelled distributions of indoor gamma-ray dose-rates

Table 2 compares parameters reflecting the goodness of fit of the four main models that have been employed and of CD averages based on the National Survey set of 2,283 measurements and on the full set of 10,199 measurements. A steady improvement in fitting can be seen from the National Survey CD mean as used in the published study of childhood cancer and natural background radiation (Kendall *et al*, 2013). Appendix G summarises the approaches of other workers to the estimation of indoor gamma-ray dose-rates.

It is clear from Table 2 that the E-OLS model provides a better fit than either of the MRGP-8 models or the Gaussian-Matérn model. Given this, and that it did not prove possible to estimate uncertainties

for the two MRGP models due to extensive numerical difficulties, it is the E-OLS model that is preferable for use in epidemiology. However, the Gaussian-Matérn model has marked theoretical advantages, and errors can be more readily propagated to dose-rate estimates. It should be noted that the addition of the extra parameters to the E-OLS model, compared to the broadly similar model reported by Kendall and co-workers (Kendall *et al*, 2016b) has resulted in a considerable improvement in the prediction accuracy (the mean square error, MSE, having dropped from 377.6 to 355.7). In a similar way, use of a much larger set of variables, some overlapping with those of the E-OLS model (Appendix C, Appendix B Table B2), has notably improved the performance of the Gaussian-Matérn model over the version used by Kendall and co-workers (Kendall *et al*, 2016b), with MSE decreasing from 410.92 to 401.74. Nevertheless, it should be noted that the Gaussian-Matérn model gives a less accurate prediction of the mean of the distribution than any of the other models; in contrast, the standard deviation is the lowest. Residuals for the Gaussian-Matérn model are compared with those of the EOLS model in Appendix C, and it will be seen that the shorter tails of the Gaussian-Matérn distribution, particularly the shorter upper tails, are largely responsible for the lower standard deviation.

However, despite the detailed modelling a substantial residual MSE remains with all of the models considered. This is because of the significant residual inter-house variation, which places an irreducible lower limit on the accuracy of such predictions. This lower limit was estimated to have a standard deviation not far below about 18 nGy/h, the magnitude of the residual error in the E-OLS model. These inter-house variations are around 20% of the mean and thus larger than the measurement error (~5%). More detailed information, particularly on the building materials used in particular dwellings, might well allow the residual variation to be reduced, if such information were to be available.

It has been argued above that the spatially-averaged distributions of dose-rates shown in Fig 2 are a better representation of the underlying distribution of indoor gamma-ray dose-rates than the raw measurement readings of Fig 1. These averaged distributions are somewhat less dispersed than that of the original measurements because individual outlying measurements are subsumed within the areal averages. This is also apparent in the distributions of predicted dose-rates at the measurement locations (Fig 4) using the E-OLS, Gaussian-Matérn and MRGP-8 models. The close agreement of these three quite different models with the CD averages is most striking and provides strong evidence that the bimodality is not an artefact of the model fitting.

Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) remark that the MRGP-8 model predictions appear to be somewhat over-smoothed. One possible reason for this is that models considered by Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) do not take into account any potential non-stationarity (spatial structure changing as a function of the coordinates) or anisotropy (directionality of spatial dependence). However, computational challenges inherent to working with a dataset of this size preclude estimation of existent non-stationary anisotropic models. Thus, the simplifying assumptions of isotropy and stationarity cannot be formally tested. Further methodological developments to the MRGP model may offer the most promising approach to accounting for these features in the spatial distribution of indoor dose-rates.

One notable feature of attempts by the present workers to model the distribution of indoor gamma-ray dose-rates has been the relatively disappointing performance of geostatistical models, although it should be noted that the performance of the classical Gaussian-Matérn model has markedly improved because of use of an enlarged set of explanatory variables. One reason for the relatively poor performance of geostatistical models may be that, with current methodology and computer technology, there are serious problems in fitting datasets as large as that used in the present study (Heaton *et al*, 2017). A further factor which is likely to affect all modelling, including the E-OLS model, is that the available explanatory variables do not include those that would probably have the greatest predictive power. These would almost certainly include details of the construction of the dwelling and of the building materials used, in particular their radioactive content. In some countries a building registry will include some information on construction (e.g., single-family home or apartment), but it is unlikely that details of radioactivity in the materials from which specific buildings are constructed will ever be available on a wide-scale – it would be simpler to measure dose-rates in the completed dwelling. It has been noted here that geology may be one of the most helpful of the available explanatory variables, but it is clearly not a “magic bullet”; Table 3 indicates that the 50K-BEDSUP variable accounted for about 8% of the total variation in the dose-rate data. French researchers (Warnery *et al*, 2015) attempting to model indoor gamma-ray dose-rates in dental clinics and veterinary practices first tried a simple kriging approach and then co-kriging with uranium potential data; the reduction in MSE was modest, from 409 to 407 (nSv/h)² on a mean of 76 nSv/h.

Conclusions

The unimodal distribution of indoor gamma-ray dose-rate measurements found in Great Britain is composed of data displaying notable variation between buildings that are broadly in the same locality. When the measurements are averaged over relatively small areas this reduces the effect of inter-house variation and measurement errors, which allows an underlying dose-rate distribution that is bimodal, if not more complex, to become apparent. The finding of a bimodal distribution of indoor gamma-ray dose-rates in Great Britain has not, to our knowledge, been previously reported. However, persuasive evidence is provided by the close similarity between the distributions obtained by areally-averaging measurement results and those obtained from three entirely different modelling procedures (Fig 4). Gamma radiation dose-rate data from other countries also show distributions that generally display bimodality (or multimodality), generated by a complex variety of factors influencing indoor dose-rates (Appendix E).

This bimodality is a general reflection of the distribution of measurements by geology, but dose-rates are also determined by other (potentially correlated) factors, such as socio-economic status and period of construction. Neither geology nor socio-economic status offers a complete explanation for the shape of the distribution, although the former may be somewhat more helpful in this respect. It is likely that models with greater predictive power could be developed if data on house construction, including the radionuclide content of the various building materials, were available.

The best predictions of indoor gamma-ray dose rates were made, amongst the models tested, by the E-OLS model – a linear combination of simple models: averages over small areas, over geologies and other parameters or using a weighted sum of measurements at neighbouring points (Appendix B). However, with the addition of selected covariates a standard geostatistical model (Appendix C) performed sufficiently well to be worth using in epidemiological analyses. The models investigated here show significant improvements over those reported previously (Kendall *et al*, 2016b). However, with the data available, a substantial inter-house variation could not be explained.

Acknowledgements The authors are grateful to Kathryn Bunch, Graham Smith, Hans Vanmarcke, Kaare Ulbak, Juhani Lahtinen, Francesco Bochicchio, Marta Garcia-Talavera, Bernd Grosche, David Pawel and the two referees for detailed and helpful comments and information on national data. Cristina Nuccetelli and Rosabianca Trevisi kindly provided data from their database of NORM in building materials. The authors are also very grateful to Jill Simpson of the University of York and to

the other UKCCS investigators for making available the results of the indoor gamma ray measurements made for the United Kingdom Childhood Cancer Study and for advice on the interpretation of the data. They are grateful to Phil Gilvin, Luke Hager and Rick Tanner at Public Health England (PHE) for advice on the dosimetry of the National Survey and the UKCCS. The Digital Land Utilisation Survey 1933-1949 (AfA213) was used under licence from the Environment Agency (The Land Utilisation Survey of Britain, 1933-1949, copyright Audrey N. Clark). J D Appleton publishes with permission from the Executive Director of the BGS. This work was supported by Children with Cancer (UK) and by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute, Division of Cancer Epidemiology and Genetics.

Appendix A: The influence of period of construction and dwelling type on indoor gamma-ray dose-rates.

This Appendix discusses information on period of construction and building type on indoor gamma-ray dose-rates. Two sources of information are available.

- The National Survey of the exposure of the UK population to naturally occurring radiation indoors (Wrixon *et al*, 1988) provides reasonably detailed information, but only for a subset of the measurement locations. The corresponding information for general locations, in particular the birth locations of study subjects in epidemiological studies, is not available.
- An approximate division into dwellings probably constructed before and after 1940 can be deduced from the Dudley Stamp land use maps developed in the 1930s. This is outlined in “Materials and Methods” and more details are given here.

Information from the National Survey

Relevant data were collected as part of the National Survey of Natural Radiation Exposures in UK Dwellings (Kendall *et al*, 2016a; Kendall *et al*, 2016b; Wrixon *et al*, 1988). Data were reported by Wrixon and co-workers (Wrixon *et al*, 1988) for 2,283 dwellings in Great Britain for which an indoor gamma-ray dose-rate measurement was available. For most of these a radon measurement was also available. Mean indoor gamma radiation dose-rates and mean indoor radon concentrations for these 2,048 dwellings are summarised in Table A1. On average, radon levels are higher in larger houses while gamma-ray dose-rates are somewhat lower. Reasons for this are not known with certainty, but the lower mean radon concentration in flats (apartments) may be a consequence of many flats being above the ground floor and thus being less liable to radon ingress from the ground.

For 280 of the 2,283 dwellings with a gamma-ray measurement, dwelling type and/or period of construction (generally the latter) were not known. The breakdown of the remaining 2,003 dwellings by dwelling type and period of construction is given in Table A2. Table A3 shows the variation in mean indoor gamma-ray dose-rate with dwelling type and period of construction.

Figure A1 gives the variation with period of construction of the indoor gamma-ray dose-rate for each type of dwelling (i.e., the data of Table A3 plotted as a figure). It may be noted that mean indoor gamma-ray dose-rates differ between dwelling types and that the proportion of dwellings of each dwelling type varies significantly from period to period. The variation with time of the mean indoor gamma-ray dose-rate for all dwellings taken together is thus hard to interpret.

For all dwelling types indoor gamma-ray dose-rates in homes constructed during 1945-64 were lower than for those homes constructed in the previous period. This decline continued during 1965-1976. For semi-detached/end-terrace houses and for flats this decline was reversed in the next period (homes constructed after 1976); for detached and mid-terrace houses the decline continued.

Table A4 gives the variation in indoor radon concentration with building type and age. Radon levels in most dwellings tend to decrease over time, while for flats they increase.

Approximate period of construction of dwellings from land use maps

Methods for constructing dwellings have changed with time. Originally walls consisted of a single layer of masonry. Then cavity walls with two layers, usually of brick, were introduced. Later the inner layer of brick was replaced by one of blocks. These changes will have affected the gamma-ray dose-rate within the dwelling. In particular, the change from cavity walls consisting of outer facing bricks with an inner layer of common bricks (which became frequent between the First and Second World War) to walls consisting of an outer layer of facing bricks with an inner layer of blocks (which was introduced after the Second World War) (National House Building Council (NHBC) Foundation 2015) broadly coincided with a reduction in indoor gamma radiation dose-rates.

Unfortunately, information on period of construction is not generally available for British dwellings. However, an approximate classification of those built before and after about 1940 can be obtained by discovering whether the house in question was in an area categorized as “Urban” or “Suburban” at this time. Those that were may be taken as probably already in existence when the map was compiled. Those built on areas not categorized as (Sub)Urban may be taken to have been built later (Appleton & Cave, 2018).

The maps available for this categorization were compiled under the direction of Laurence Dudley Stamp in the 1930s (Southall *et al.* 2007). Fourteen categories of land use were assigned. Maps for England and Wales (and a small part of Southern Scotland) were converted for use in a GIS in a project of the Environment Agency about ten years ago (Southall *et al.* 2007) (Environment Agency 2010). This was a substantial task, partly because the maps had been printed over a long period by different printers and colour codings were inconsistent between sheets, but also because details such as text and contour lines were printed in colours also used for land use categories. For the present study, these GIS maps for England and Wales were used.

No GIS-ready versions of the Dudley Stamp land use maps for Scotland were available. Nine sheets covering areas of the highest population density were selected (see Table A5) and polygonised for GIS use. While much of the Scottish land area was not included in this exercise, a total of about 78% of the Scottish population were included and, in total, 98% of the population of Great Britain. Overall, 56% and 44% of the GB dwellings were classified as constructed pre- and post-1940, respectively. In Scotland and Wales, the proportions were more nearly equal than in GB as a whole (50% and 47% constructed pre-1940 for Wales and Scotland, respectively; these figures exclude the 22% of Scottish dwellings that were unclassified).

It is clear that this Pre/Post-1940 construction classification can only be approximate. Modern buildings which were built in pre-1940 areas during infill or redevelopment will be wrongly classified as “pre-1940”. Testing the predictions is difficult because they relate to a spectrum of years from 1962 until 2010. The acid test in the present context is how useful this parameter proves as a predictive tool. Nevertheless, approximate testing may be of interest.

The Department for Communities and Local Government published an English Housing Survey Housing stock report for 2008 (Department for Communities and Local Government, DCLG, 2010) in which it was reported (their Table 1.1) that 43% of English housing was built before 1945; the total for Great Britain will be similar. These figures exclude the “Post-1990” category which covers only the end of the period used in the present work. As described above, the UK National Survey of Natural Radiation Exposure in UK Dwellings (Wrixon *et al.* 1988) collected data on period of construction. The percentages of the sample built before and after 1944 were 47% and 53%, respectively. Both the DCLG and National Survey data thus have a slight majority of dwellings constructed after 1944 while our Dudley Stamp data indicates a similar sized majority pre-1940. It is unlikely that the difference between 1944 and 1940 as the cut-off is significant. As noted above, the Dudley Stamp classification

will fail to pick up redevelopment or infill and this probably accounts for the difference. Despite its approximate nature the Dudley Stamp pre- and post-1940 construction classification is found to have predictive value for indoor gamma-ray dose-rate modelling. Figure A2 compares the median indoor gamma-ray dose-rates for dwellings constructed pre- and post-1940 for different underlying geologies. In each case the dose-rate in the newer houses is lower than that in older ones indicating that the age of construction effect is not a consequence of changes in location.

Table A1 Number of measurements with mean indoor radon level and gamma-ray dose-rate by house type based on the data of the UK National Survey.

Type of Residence	Number of dwellings	Mean radon concentration Bq/m ³	Mean gamma-ray dose-rate nGy/h
Detached House	511	33.2	89.7
Semi-detached or end-terrace house	807	19.3	98.7
Mid-terrace house	459	20.7	97.4
Flat/Maisonette	250	13.9	85.9
Unknown type	21	13.7	76.4
Total	2,048	22.4	94.4

Table A2 Numbers of dwellings by dwelling type and period of construction (with percentages)

	<i>Before 1900</i>	<i>1900-19</i>	<i>1920-44</i>	<i>1945-64</i>	<i>1965-76</i>	<i>After 1976</i>	<i>Total</i>
Detached House	83 (31.8)	28 (13.6)	92 (19.0)	122 (24.4)	133 (32.4)	55 (38.5)	513
Semi-detached or end-terrace house	56 (21.5)	73 (35.4)	266 (55.1)	240(48.1)	147(35.8)	31 (21.7)	813
Mid-terrace house	97 (37.2)	86 (41.7)	103 (21.3)	69 (13.8)	69(16.8)	27 (18.9)	451
Flat/Maisonette	25 (37.2)	19 (9.2))	22 (4.6)	68 (13.6)	62 (15.1)	30 (21.0)	226
Total	261 (100)	206 (100)	483 (100)	499 (100)	411 (100)	143 (100)	2,003

Table A3 Mean indoor gamma-ray dose-rate (nGy/h) by dwelling type and period of construction

	<i>Before 1900</i>	<i>1900-19</i>	<i>1920-44</i>	<i>1945-64</i>	<i>1965-76</i>	<i>After 1976</i>	Total
Detached House	89.0	105.2	95.8	89.6	85.5	82.1	89.6
Semi-detached or end-terrace house	94.1	101.7	102.7	100.3	88.8	95.0	98.5
Mid-terrace house	95.8	100.7	104.5	99.1	85.8	80.5	96.8
Flat/Maisonette	85.9	86.6	98.4	87.0	77.7	83.9	85.0
Total	92.3	100.4	101.6	95.7	85.5	85.0	94.3

Table A4 Mean indoor radon concentrations (Bq.m⁻³) by dwelling type and period of construction

	Before 1900	1900-19	1920-44	1945-64	1965-76	After 1976	Total
Detached House	28.6	49.8	24.0	34.8	23.1	23.4	33.2
Semi-detached or end-terrace house	24.5	15.7	16.9	18.6	17.4	19.8	19.3
Mid-terrace house	20.6	29.2	14.0	17.5	17.8	13.7	20.7
Flat/Maisonette	13.1	9.0	9.2	11.6	13.9	20.1	13.9
Total	25.7	28.4	18.4	22.6	19.9	21.6	22.4

Table A5 Sheets of the Dudley Stamp Land Use maps for Scotland that were converted for GIS use

<i>Sheet</i>	<i>Name</i>
58	Arbroath and Montrose
64	Dundee and St Andrews
66	Loch Lomond
67	Stirling and Dunfermline
68	Firth of Forth
72	Glasgow
73	Falkirk and Motherwell
74	Edinburgh
78	Kilmarnock and Ayr

Fig A1 Breakdown of mean indoor gamma-ray dose-rates by dwelling type and year of construction

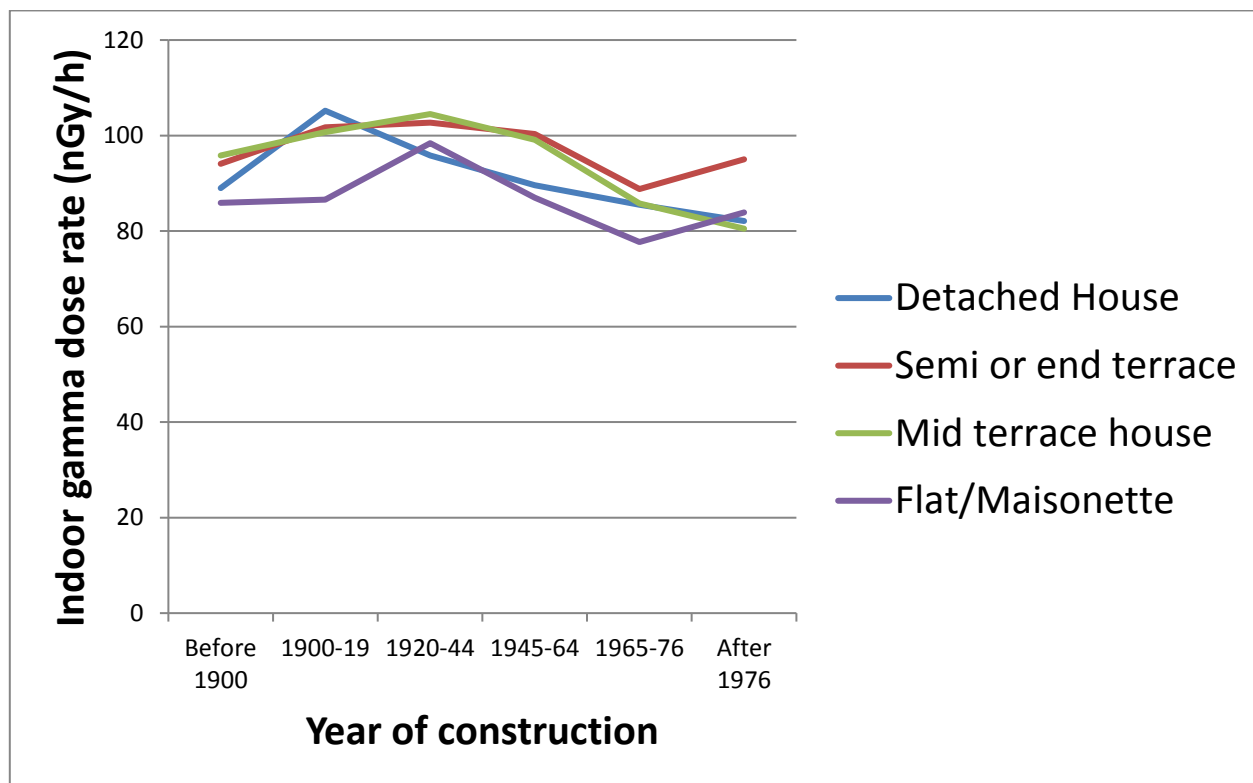
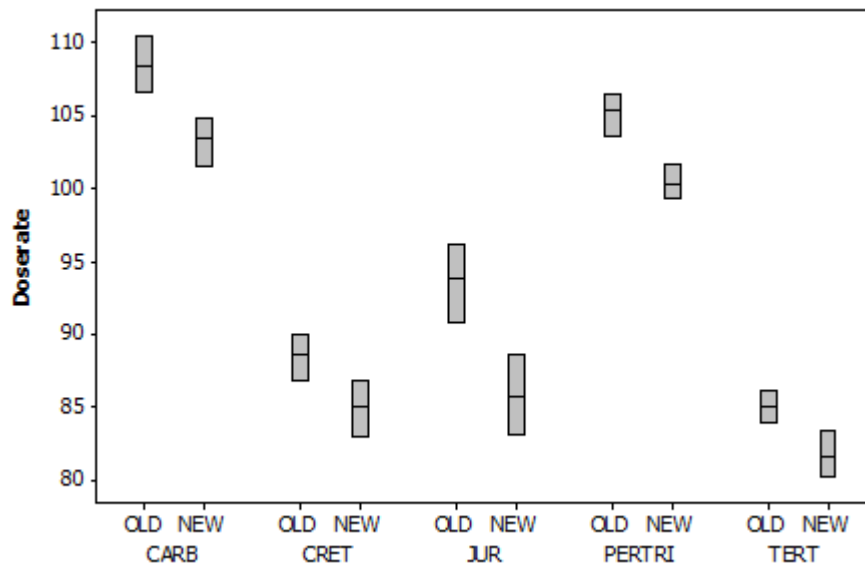


Fig A2 Median indoor gamma-ray dose-rates (nGy/hr) for dwellings constructed pre-1940 (“OLD”) compared to those constructed post-1940 (“NEW”) by underlying geology (Carboniferous, Cretaceous, Jurassic, Permian-Triassic and Tertiary). (Box = median 95% confidence limits; Horizontal line in box = median; Pre-1940 data derived from © L. Dudley Stamp/Geographical Publications Ltd, Audrey N. Clark, Environment Agency/DEFRA and Great Britain Historical GIS)



Appendix B.: Revised ordinary least squares linear-regression analysis

Indoor measurement data and other covariates used

The indoor gamma-ray dose-rate measurement data are as reported previously (Kendall *et al*, 2016b). There have been a few small changes made to coding of some of the other data, in particular the 50k-BEDSUP-Surface codes, county district, Carstairs score and population density (see OLR-1). Here the candidate models were refitted using a number of candidate dose-interpolation measures. These comprise jackknife estimates (Davison & Hinkley, 1997) of dose-rate based on interpolated “nearby” dose-rate measurements. Three basic types of interpolation were used: (a) weighting of neighbouring

dose-rates by an inverse power of distance; (b) an average over some administratively defined areas (e.g., county district); or (c) an average over some geologically defined areas.

There are small differences in the present analysis from what was done in the previous paper (Kendall *et al*, 2016b) relating to the simple areal-average estimates, types (b) and (c) above. Previously, the jackknife area means, e.g., for 50k-BEDSUP, were computed for each candidate point i by finding the nearest point (using Euclidean distance) in the remaining $n = 10,198$ dataset, $j_{\min}(i)$, then using the mean dose-rate of all points having the same, e.g., 50k-BEDSUP, value as $j_{\min}(i)$ in the $n = 10,198$ data and assigning that to the candidate point. In the present paper area means for the relevant area are simply averaged, e.g., given by 50k-BEDSUP code, corresponding to the candidate point i using the mean dose-rate of all points having the same, e.g., 50k-BEDSUP, value as i in the $n = 10,198$ data.

A number of supplementary geological measures are also included in the present analysis, specifically a 23-level 50k-BEDSUP-surface code, a 23-level 50k-BEDSUP-surface-bedrock code and a 3-level (pre-1940, post-1940, unknown) house construction period code that were not previously employed. The variables used are listed in Table B1. The optimal model and associated parameter values are given in Table B2. As can be seen, the model is somewhat different in form from the previously fitted optimal model (Table 17 in (Kendall *et al*, 2016b)).

Empirical ordinary least squares model selection and fitting, and Gaussian process maximum likelihood fitting

As previously, a highly parameterized empirical model was constructed based on linear combinations of the interpolation measures described in the previous section. In order to construct an empirical model that satisfactorily explains the spatial variation of mean dose-rate, linear regression with ordinary least squares (OLS) (Rao, 2002) was used. This model assumes a standard Normal error, and attempts to model the spatial correlation in dose-rate using combinations of explanatory variables. In order to avoid over-parameterised models, the Akaike Information Criterion (AIC) (Akaike, 1973; Akaike, 1981) was minimized to select the optimal set of interpolation variables (the calculation of which is outlined above). AIC penalises against overfitting by adding $2 \times$ [number of fitted parameters] to the model deviance (residual sum of squares). An iterative mixed-forward-backward stepwise procedure was used to minimise AIC for OLS using R (R Project version 3.2.2, 2015). There is literature, e.g. (Hurvich & Tsai, 1989), suggesting that the AIC may lead in some cases (in a class of

autoregressive models) to over-parameterised models where the datasets are small, but not for problems such as the present ones.

The fit of all these models was tested using a standard cross-validation process. The models were fitted to a randomly selected 70% of the data, and the indicated models were then used to predict gamma-ray dose-rates in the remaining 30% of the data. The randomly chosen 70% (and 30%) samples were identical to those previously used (Kendall *et al*, 2016b). The resulting mean-square error was 355.71 (Table 2), a substantial decrease on the previously reported estimate, 377.64 (Kendall *et al*, 2016b). It is possible that another random 70-30 partition of the data would result in a different optimal model with a different MSE. Here, a different partition was not explored because: 1) there are currently no guidelines as to how many such random partitions one must explore; and 2) a reliable method to combine predictions across multiple random partitions does not exist in a frequentist setting. In a Bayesian setting, one could employ Bayesian model averaging, but this falls outside the scope of the current paper.

Calculating uncertainties in E-OLS estimates

A possible way of estimating the errors in estimated dose-rates taking account of the indicated non-normality of model residuals, is as follows:

1) Take N bootstrap samples from the $M=10,199$ measurement data and estimate the projected dose-rates for the $K = 124,454$ measurement data points for each, resulting in a $K \times N$ bootstrap array

2) For each of $n = 1, \dots, K$, estimate percentile confidence intervals (CIs) in the usual way from this $K \times N$ array, sampling for each $k=1, \dots, K$ the percentile-based CI based on the N dose-rate estimates.

Bias corrected adjusted bootstrap CIs (Efron, 1987) could in principle be calculated in step (2), but the computational overheads might be very onerous.

Table B1 Variables used in ordinary least squares (OLS) regressions as part of the forward-backwards stepwise Akaike Information Criterion (AIC) variable selection.

Variable
Northing
Easting
Northing ²
Easting ²
Grouping into 16 geological groups <i>a priori</i> by 50K-BEDSUP Bedrock codes
Grouping into 23 geological groups <i>a priori</i> by 50K-BEDSUP Surface codes
Grouping into 23 geological groups <i>a priori</i> by combined 50K-BEDSUP Surface-Bedrock codes
Average gamma-ray dose-rate by 50K-BEDSUP (25 groups)
Average gamma-ray dose-rate by 50K-BEDSUP (50 groups)
Average gamma-ray dose-rate by 50K-BEDSUP (100 groups)
Average gamma-ray dose-rate by 50K-BEDSUP (200 groups)
Average gamma-ray dose-rate by county district
Average gamma-ray dose-rate by Post Code District
Average gamma-ray dose-rate by 10 km square
Average gamma-ray dose-rate by 50K-BEDSUP
Average gamma-ray dose-rate by LEX-ROCK
Average gamma-ray dose-rate by LEX-ROCK-Lex
Average gamma-ray dose-rate by LEX-ROCK-Rock
Average gamma-ray dose-rate by 50K-BEDSUP-Bedrock
Average gamma-ray dose-rate by 50K-BEDSUP-Surface
Average gamma-ray dose-rate weighted by 1/distance ^{0.0}
Average gamma-ray dose-rate weighted by 1/distance ^{0.5}
Average gamma-ray dose-rate weighted by 1/distance ^{1.0}
Average gamma-ray dose-rate weighted by 1/distance ^{1.5}
Average gamma-ray dose-rate weighted by 1/distance ^{2.0}
Average gamma-ray dose-rate weighted by 1/distance ^{2.5}
Average gamma-ray dose-rate weighted by 1/distance ^{3.0}
Average gamma-ray dose-rate weighted by 1/distance ^{0.5} (different geology), 1/distance ^{1.0} (same geology)
Average gamma-ray dose-rate weighted by 1/distance ^{0.5} (different geology), 1/distance ^{1.5} (same geology)
Average gamma-ray dose-rate weighted by 1/distance ^{1.0} (different geology), 1/distance ^{0.5} (same geology)
Average gamma-ray dose-rate weighted by 1/distance ^{1.0} (different geology), 1/distance ^{1.5} (same geology)
Average gamma-ray dose-rate weighted by 1/distance ^{1.5} (different geology), 1/distance ^{0.5} (same geology)
Average gamma-ray dose-rate weighted by 1/distance ^{1.5} (different geology), 1/distance ^{1.0} (same geology)
Country code (England, Wales, Scotland)
Pre- vs post-1940 classification (yes, no, unknown)
Carstairs score (1981)
Population density (1981)
Urban-rural group (1981) [6-level]

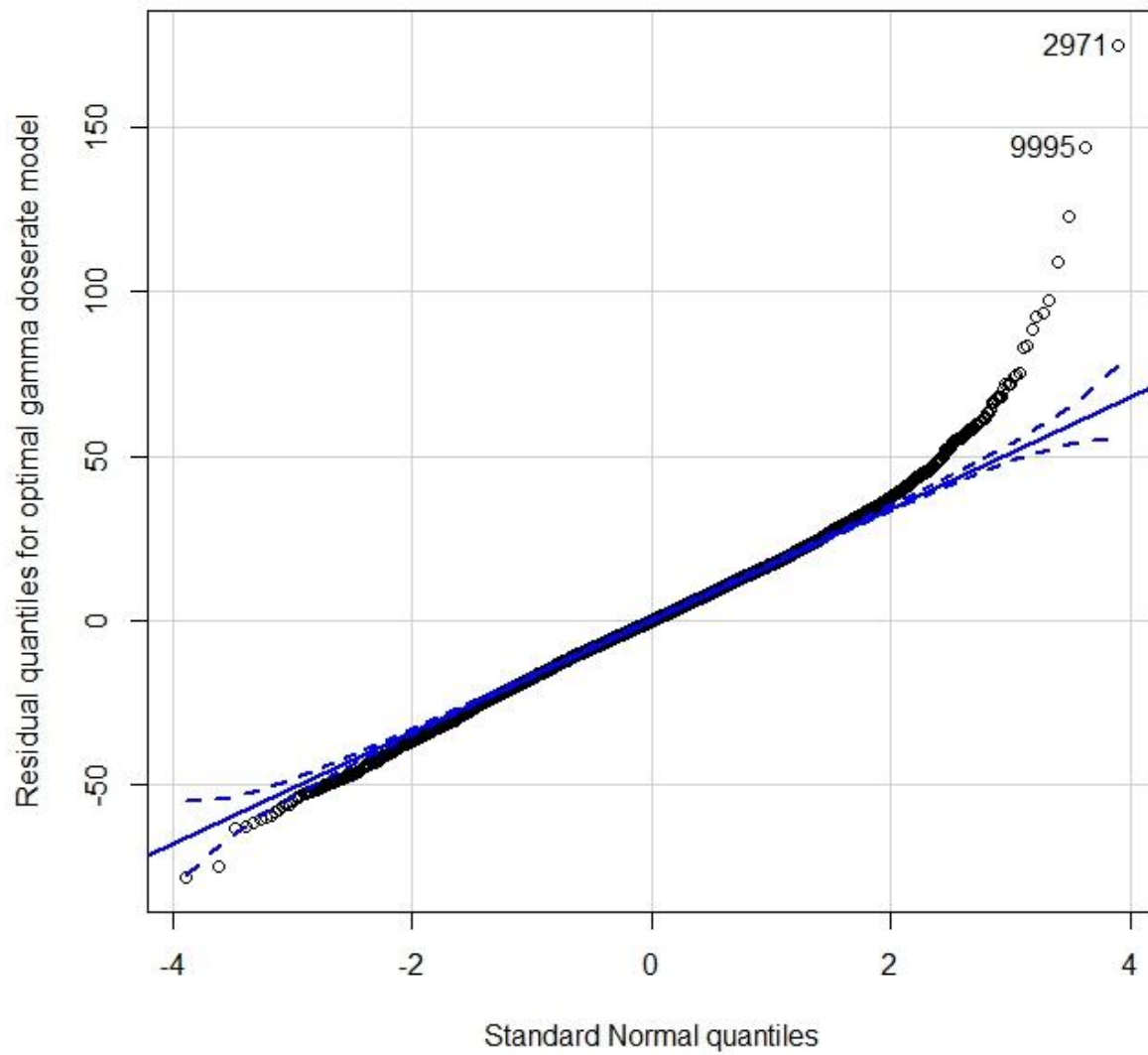
Table B2 Parameter values and 95% CIs for AIC-optimal model fitted by ordinary least squares. Models are listed in the order they entered the AIC-optimal model.^a

Parameter	Value (+95% CI)
Constant (Intercept)	254.994 (222.912, 302.511)
Average gamma-ray dose-rate weighted by 1/distance ^{1,0}	0.235 (-0.076, 0.547)
Average gamma-ray dose-rate by 50K-BEDSUP group	-1.382 (-1.628, -1.213)
Unknown vs post-1940 classification	2.593 (0.322, 5.030)
Pre-1940 vs post-1940 classification	2.938 (2.157, 3.685)
Average gamma-ray dose-rate by 50K-BEDSUP (group 2 vs group 1)	-31.623 (-54.157, -12.380)
Average gamma-ray dose-rate by 50K-BEDSUP (group 3 vs group 1)	-30.220 (-49.822, -12.908)
Average gamma-ray dose-rate by 50K-BEDSUP (group 4 vs group 1)	-51.965 (-73.835, -34.366)
Average gamma-ray dose-rate by 50K-BEDSUP (group 5 vs group 1)	-51.903 (-70.807, -36.438)
Average gamma-ray dose-rate by 50K-BEDSUP (group 6 vs group 1)	-59.371 (-78.695, -43.416)
Average gamma-ray dose-rate by 50K-BEDSUP (group 7 vs group 1)	-61.158 (-80.129, -45.760)
Average gamma-ray dose-rate by 50K-BEDSUP (group 8 vs group 1)	-64.321 (-83.679, -48.872)
Average gamma-ray dose-rate by 50K-BEDSUP (group 9 vs group 1)	-66.501 (-85.870, -50.893)
Average gamma-ray dose-rate by 50K-BEDSUP (group 10 vs group 1)	-67.213 (-86.883, -51.547)
Average gamma-ray dose-rate by 50K-BEDSUP (group 11 vs group 1)	-70.770 (-90.303, -55.186)
Average gamma-ray dose-rate by 50K-BEDSUP (group 12 vs group 1)	-72.614 (-92.140, -57.057)
Average gamma-ray dose-rate by 50K-BEDSUP (group 13 vs group 1)	-73.172 (-93.101, -57.447)
Average gamma-ray dose-rate by 50K-BEDSUP (group 14 vs group 1)	-76.828 (-96.768, -61.485)
Average gamma-ray dose-rate by 50K-BEDSUP (group 15 vs group 1)	-76.167 (-96.184, -60.442)
Average gamma-ray dose-rate by 50K-BEDSUP (group 16 vs group 1)	-79.335 (-99.289, -63.588)
Average gamma-ray dose-rate by 50K-BEDSUP (group 17 vs group 1)	-79.611 (-100.020, -63.992)
Average gamma-ray dose-rate by 50K-BEDSUP (group 18 vs group 1)	-81.069 (-101.402, -65.264)
Average gamma-ray dose-rate by 50K-BEDSUP (group 19 vs group 1)	-82.994 (-103.232, -67.218)
Average gamma-ray dose-rate by 50K-BEDSUP (group 20 vs group 1)	-85.079 (-105.287, -68.969)
Average gamma-ray dose-rate by 50K-BEDSUP (group 21 vs group 1)	-84.495 (-105.252, -68.730)
Average gamma-ray dose-rate by 50K-BEDSUP (group 22 vs group 1)	-86.658 (-107.776, -70.727)
Average gamma-ray dose-rate by 50K-BEDSUP (group 23 vs group 1)	-90.814 (-112.672, -74.667)
Average gamma-ray dose-rate by 50K-BEDSUP (group 24 vs group 1)	-93.919 (-115.478, -77.826)
Average gamma-ray dose-rate by 50K-BEDSUP (group 25 vs group 1)	-96.147 (-117.781, -79.784)
Average gamma-ray dose-rate by 50K-BEDSUP (group 26 vs group 1)	-100.187 (-122.386, -83.335)
Average gamma-ray dose-rate by 50K-BEDSUP (group 27 vs group 1)	-95.983 (-118.259, -79.103)
Average gamma-ray dose-rate by 50K-BEDSUP (group 28 vs group 1)	-100.640 (-122.621, -84.629)
Average gamma-ray dose-rate by 50K-BEDSUP (group 29 vs group 1)	-102.515 (-124.375, -86.449)
Average gamma-ray dose-rate by 50K-BEDSUP (group 30 vs group 1)	-100.683 (-123.288, -83.871)
Average gamma-ray dose-rate by 50K-BEDSUP (group 31 vs group 1)	-103.617 (-126.106, -87.129)
Average gamma-ray dose-rate by 50K-BEDSUP (group 32 vs group 1)	-104.674 (-126.705, -88.422)
Average gamma-ray dose-rate by 50K-BEDSUP (group 33 vs group 1)	-106.431 (-128.984, -90.037)
Average gamma-ray dose-rate by 50K-BEDSUP (group 34 vs group 1)	-109.940 (-132.475, -93.578)
Average gamma-ray dose-rate by 50K-BEDSUP (group 35 vs group 1)	-111.499 (-134.128, -95.227)
Average gamma-ray dose-rate by 50K-BEDSUP (group 36 vs group 1)	-112.078 (-134.827, -95.455)
Average gamma-ray dose-rate by 50K-BEDSUP (group 37 vs group 1)	-113.663 (-136.430, -97.094)

Average gamma-ray dose-rate by 50K-BEDSUP (group 38 vs group 1)	-113.825 (-136.770, -97.159)
Average gamma-ray dose-rate by 50K-BEDSUP (group 39 vs group 1)	-115.847 (-138.951, -99.327)
Average gamma-ray dose-rate by 50K-BEDSUP (group 40 vs group 1)	-118.400 (-142.244, -101.491)
Average gamma-ray dose-rate by 50K-BEDSUP (group 41 vs group 1)	-119.562 (-143.163, -102.834)
Average gamma-ray dose-rate by 50K-BEDSUP (group 42 vs group 1)	-119.778 (-143.442, -103.033)
Average gamma-ray dose-rate by 50K-BEDSUP (group 43 vs group 1)	-119.885 (-143.692, -102.779)
Average gamma-ray dose-rate by 50K-BEDSUP (group 44 vs group 1)	-124.763 (-148.818, -107.499)
Average gamma-ray dose-rate by 50K-BEDSUP (group 45 vs group 1)	-129.758 (-154.046, -112.073)
Average gamma-ray dose-rate by 50K-BEDSUP (group 46 vs group 1)	-132.794 (-157.646, -114.762)
Average gamma-ray dose-rate by 50K-BEDSUP (group 47 vs group 1)	-135.816 (-161.687, -117.934)
Average gamma-ray dose-rate by 50K-BEDSUP (group 48 vs group 1)	-142.463 (-168.376, -124.569)
Average gamma-ray dose-rate by 50K-BEDSUP (group 49 vs group 1)	-138.572 (-166.860, -118.873)
Average gamma-ray dose-rate by 50K-BEDSUP (group 50 vs group 1)	-152.503 (-187.913, -126.754)
Urban/rural 1981 area 2 vs area 1	-1.820 (-2.811, -0.864)
Urban/rural 1981 area 3 vs area 1	-0.410 (-1.876, 1.073)
Urban/rural 1981 area 4 vs area 1	-0.418 (-2.368, 1.424)
Urban/rural 1981 area 5 vs area 1	-5.186 (-8.110, -2.239)
Urban/rural 1981 area 6 vs area 1	-2.524 (-4.026, -1.039)
Easting (km)	-0.01237 (-0.04680, 0.02181)
Easting (km) ² x 10 ⁶	25.058 (-15.661, 65.668)
County district average	0.122 (0.039, 0.205)
Average gamma-ray dose-rate by LEX-ROCK bedrock group	0.133 (0.006, 0.255)
Average gamma-ray dose-rate weighted by 1/distance ^{2.5}	0.068 (0.004, 0.133)
Average gamma-ray dose-rate by 50K-BEDSUP surface group	-0.076 (-0.174, 0.016)
Average gamma-ray dose-rate weighted by 1/distance ^{0.5} (different geology), 1/distance ^{1.0} (same geology)	0.224 (-0.047, 0.489)
Average gamma-ray dose-rate by PCD group	0.034 (-0.009, 0.076)
Average gamma-ray dose-rate by 50K-BEDSUP bedrock group	-0.084 (-0.203, 0.019)

^aNote: a linear term in Easting has been added to the OLS optimal model to render it algebraically complete.

Fig B1 Normal-quantile plot of optimal model residuals (observed - model expected). The two label numbers are the sequence numbers of these points in the file.



Appendix C Gaussian-Matérn maximum likelihood fitting

Here it is outlined how a standard geostatistical model was fitted to the indoor gamma-ray dose-rate data by maximum-likelihood (Diggle & Ribeiro, 2007). Specifically, it was assumed that the $\ln[\text{gamma-ray dose-rate}] Y_i$ at spatial location x_i was given by:

$$Y_i = \mu(x_i) + S(x_i) + Z_i \quad (\text{C1})$$

where $\mu(x_i) = E[Y_i]$ is the mean of Y_i , S is a stationary Gaussian process with mean 0, variance σ^2 , and spatial correlation $\rho(u) = \text{Corr}[S(x), S(x-u)]$, and Z_i are independent Gaussian random variables with mean 0 and nugget variance τ^2 . The Matérn model (Matérn, 1960) was employed, which assumes that:

$$\rho(u | \phi, \kappa) = 2^{\kappa-1} (u / \phi)^\kappa K_\kappa(u / \phi) \quad (\text{C2})$$

where $K_\kappa(\cdot)$ is a modified Bessel function of order κ (Diggle & Ribeiro, 2007). The Matérn model was chosen because of its flexibility. The integer part of the parameter κ determines the mean-square differentiability of the process S – the process is mean-square differentiable when $\kappa > 1$, and mean-square twice differentiable when $\kappa > 2$. Formal estimation of the parameter κ is difficult and was not attempted here. Preliminary curve fitting and examination of the empirical variogram, by eye, suggested that $\kappa = 2.5$ gave a reasonable fit; this value was used in all analyses, as in previous such modelling (Kendall *et al*, 2016b). Various models of the mean process, $\mu(x)$, were constructed as outlined in Table C1. All associated model parameters ($\mu(x)$, ϕ , τ^2) were estimated by maximum likelihood.

$(Y_i)_{i=1}^N$ is multivariate Gaussian with mean $\mu(x_i)$ and variance $\sigma^2 V = \sigma^2 [R + \tau^2 I] = \sigma^2 R + \tau^2 I$ where I is the $N \times N$ identity matrix, and $R = (r_{ij})_{i,j=1}^N = (\rho(\|x_i - x_j\|))_{i,j=1}^N$. If now one has an arbitrary point x , with associated “true” (unobserved) $\ln[\text{gamma-ray dose rate}] T(x) = \mu(x) + S(x)$, then $(T(x), (Y_i)_{i=1}^N)$ is multivariate Gaussian with mean $(\mu(x), (\mu(x_i))_{i=1}^N)$ and variance $\begin{bmatrix} \sigma^2 & \sigma^2 r^T \\ \sigma^2 r & \sigma^2 V \end{bmatrix}$ where the vector r is given by $r = (r_i)_{i=1}^N = (\rho(\|x - x_i\|))_{i=1}^N$. Then it is easily shown (e.g., p.136 in

(Diggle & Ribeiro, 2007)) that the conditional distribution of $T(x)$ given $(Y(x_i))_{i=1}^N$ is also multivariate Gaussian with mean:

$$m(x) = \mu(x) + r^T V^{-1} [Y(x_i) - \mu(x_i)]_{i=1}^N \quad (C3)$$

and variance:

$$v(x) = \sigma^2 [1 - r^T V^{-1} r] \quad (C4)$$

Finally one can transform back to the natural dose-rate scale, so that the mean doserate at x is given by:

$$M(x) = \exp[m(x) + v(x) / 2] \quad (C5)$$

In estimating these the maximum-likelihood estimates of $\mu(), V, \sigma^2$, namely $\hat{\mu}(), \hat{V}, \hat{\sigma}^2$ were substituted.

The optimal model includes Easting and Northing and also the Dudley Stamp Pre/Post 1940 construction category, the urban-rural (6-level) classification, the external gamma dose rate and the 50K-BEDSUP Surface Geological classification (23-level). The fit of this model was tested using a standard cross-validation process, by fitting to the randomly selected 70% of the data, and the indicated model was then used to predict gamma-ray dose-rate in the remaining 30% of the data. Table 2 of the main text shows the mean-square-error estimated in the 30% test sample.

Table C2 compares quantiles of the distributions of measured doserates with those predicted by the optimal E-OLS and Gaussian-Matérn models. It can be seen that both models have a narrower spread than the actual data. While Gaussian-Matérn and E-OLS results are broadly similar, the E-OLS predicts a slightly longer upper tail, which accounts for the lower SD of the Gaussian-Matérn model. Figure C1 compares the distribution of residuals (i.e., of differences observed-predicted values at the measurement points) for the EOLS and for the Gaussian-Matérn models. It can be seen that while the former are symmetrically distributed about zero the latter are slightly displaced towards positive values, indicating that the Gaussian-Matérn predictions are somewhat lower.

Table C1 Log likelihood, Akaike Information Criterion (AIC) and number of fitted parameters in fits of Gaussian-Matérn model to $\ln[\text{gamma dose rate}]$. Optimal model in boldface. Unless otherwise stated all p -values relate to the improvement in fit from the immediately preceding model in the Table.

Model description/variables used	Parameters for Matérn covariance	Other model parameters	$\ln[\text{likelihood}]$	AIC	p -value
Constant	2	1	10369.92	-20733.85	
$\ln[\text{external gamma dose rate}]$	2	2	10374.95	-20741.90	0.0015
Carstairs 1981	2	2	10375.63	-20743.26	0.0007 ^a
Linear Easting+Northing	2	3	10370.46	-20730.93	0.5824 ^a
Linear-quadratic Easting+Northing + Easting x Northing	2	6	10390.58	-20765.17	<0.0001 ^a
Urban-rural (6-level)	2	6	10402.42	-20788.83	<0.0001 ^a
Dudley Stamp house construction (pre- vs post-1940)	2	3	10420.04	-20830.08	<0.0001 ^a
England-Wales-Scotland	2	3	10370.70	-20731.40	0.4594 ^a
50K-BEDSUP Bedrock (16-level)	2	16	10386.52	-20737.03	0.0044 ^a
50K-BEDSUP Surface (23-level)	2	23	10394.21	-20738.42	0.0009 ^a
50K-BEDSUP Bedrock-Surface (23-level)	2	23	10392.82	-20735.65	0.0021 ^a
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing	2	8	10438.73	-20857.47	<0.0001 ^b / <0.0001 ^c
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + urban-rural (6-level)	2	13	10461.97	-20893.94	<0.0001
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + urban-rural (6-level) + Carstairs	2	14	10462.24	-20892.49	0.4579
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + urban-rural (6-level) + $\ln[\text{external gamma dose rate}]$	2	14	10466.60	-20901.20	0.0023 ^d
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + urban-rural (6-level) + $\ln[\text{external gamma dose rate}]$ + 50K-BEDSUP Surface (23-level)	2	36	10492.50	-20909.00	0.0003
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + urban-rural (6-level) + $\ln[\text{external gamma dose rate}]$ + 50K-BEDSUP Bedrock (16-level)	2	29	10481.03	-20900.05	0.0168 ^e
Dudley Stamp + linear-quadratic Easting+Northing + Easting x Northing + urban-rural (6-level) + $\ln[\text{external gamma dose rate}]$ + 50K-BEDSUP Surface (23-level) + 50K-BEDSUP Bedrock (16-level)	2	51	10501.04	-20896.08	0.0107 / 0.3142 ^f

^a p -value of improvement in fit over model with only constant term

^b p -value of improvement in fit over model with only linear-quadratic Easting + Northing + Easting x Northing terms

^c p -value of improvement in fit over model with only Dudley Stamp house construction variable

^d p -value of improvement in fit over model with Dudley Stamp house construction, linear-quadratic Easting + Northing + Easting x Northing + urban-rural

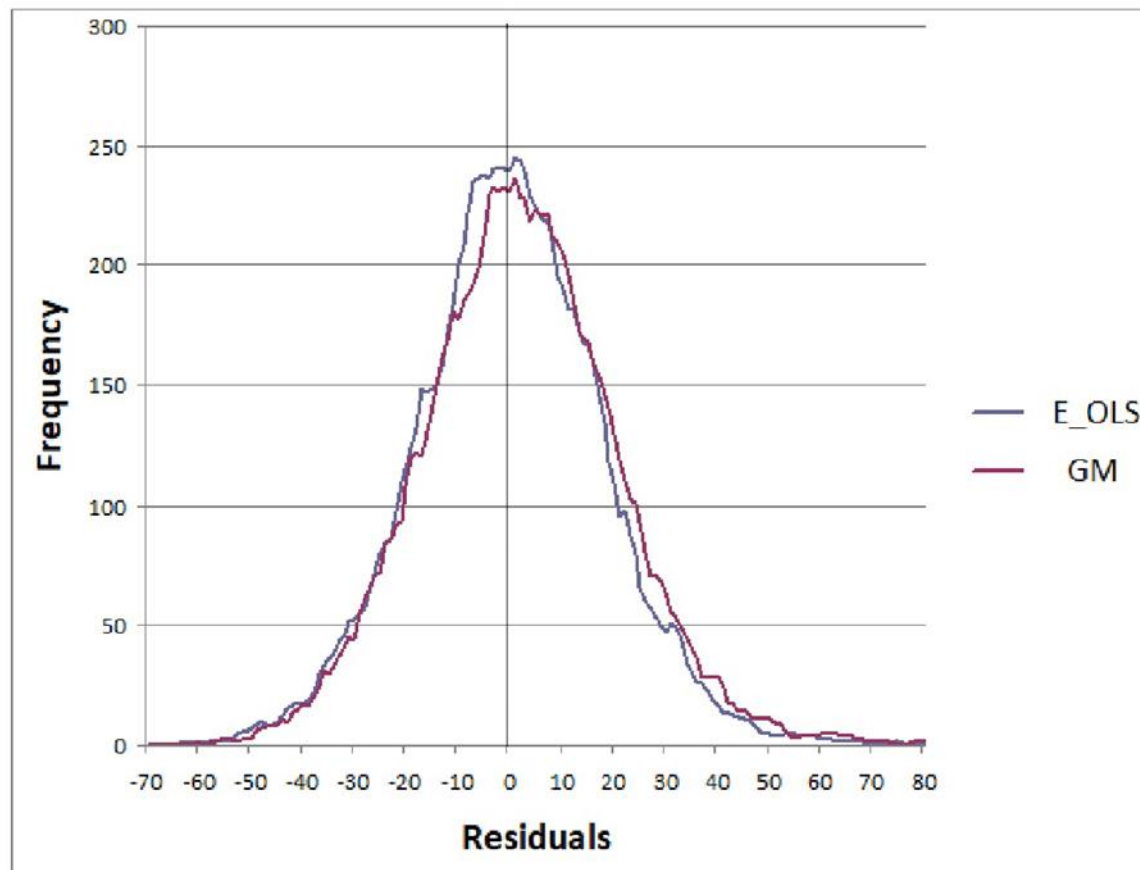
^e p -value of improvement in fit over model with Dudley Stamp house construction, linear-quadratic Easting + Northing + Easting x Northing + urban-rural + $\ln[\text{external gamma dose rate}]$

^f p -value of improvement in fit over model with Dudley Stamp house construction, linear-quadratic Easting + Northing + Easting x Northing + urban-rural + $\ln[\text{external gamma dose rate}]$ + 50K-BEDSUP Surface

Table C2_Quantiles of distribution of actual doserate and model fitted doserate for optimal E-OLS and Gaussian-Matérn model

Percentile	True doserate (nGy/h)	E-OLS model predicted doserate (nGy/h)	Gaussian-Matérn model predicted doserate (nGy/h)
1	45.64	65.58	67.47
2.5	52.00	71.23	70.92
5	58.92	75.35	75.05
10	67.10	79.76	78.41
25	80.34	85.72	84.05
50	95.32	97.11	95.68
75	110.15	105.25	102.72
90	123.87	110.73	107.92
95	132.43	113.63	110.64
97.5	140.77	116.53	112.74
99	151.95	121.77	114.18

Fig C1 Distribution of residuals for EOLS and GM Models



Appendix D. Decomposition of the observed indoor gamma-ray dose-rate distribution as a superposition of Normal random variables

A more formal decomposition of the measured indoor gamma-ray dose-rate distribution as a superposition of Normal distributions may be made as follows. It was assumed that the dose-rates were given by a weighted superposition of N Normally distributed random variables, comprising $N=3$ distributions with means, μ_i , at 47.5, 52.5, 57.5 nGy/h, $N_2=60$ with means at 60.5, 61.5, 62.5, ..., 119.5 nGy/h, then a further $N_3=17$ with means at 122.5, 127.5, 132.5, ..., 197.5, 202.5 nGy/h with probabilities p_i . The distribution at each of the $N = N_1 + N_2 + N_3 = 80$ points is $X_i \sim N(\mu_i, \sigma^2)$, with common standard deviation σ , so that the overall distribution is a superposition of these. [Note: this ensemble of random variables is not independent, since a given individual can be assumed to drawn from one and only one of the N distributions with probability p_i , so that the probability of the individual being drawn from distribution i and distribution $j \neq i$ is $0 \neq p_i p_j$.] This implies that the overall cumulative density function is:

$$F(d | (p_i)_{i=1}^N, (\mu_i)_{i=1}^N, \sigma) = P[X \leq d | (\mu_i)] = \sum_{i=1}^N p_i P[N(\mu_i, \sigma^2) \leq d] \quad (D1)$$

The probabilities $(p_i)_{i=1}^N$ are constrained so as to be positive and $\sum_{i=1}^N p_i = 1$. This cumulative distribution function was fitted to the empirical data, consisting of counts of persons in each of the $M = 80$ intervals defined by the cut-points $d_0 = 0, d_1 = 50, d_2 = 55, d_3 = 60, d_4 = 61, d_5 = 62, \dots, d_5 = 119, d_{63} = 120, d_{64} = 125, d_{66} = 130, \dots, d_{79} = 200, d_{80} = \infty$ nGy/h via (multinomial) maximum likelihood, i.e. by maximizing in the weights $(p_i)_{i=1}^N$:

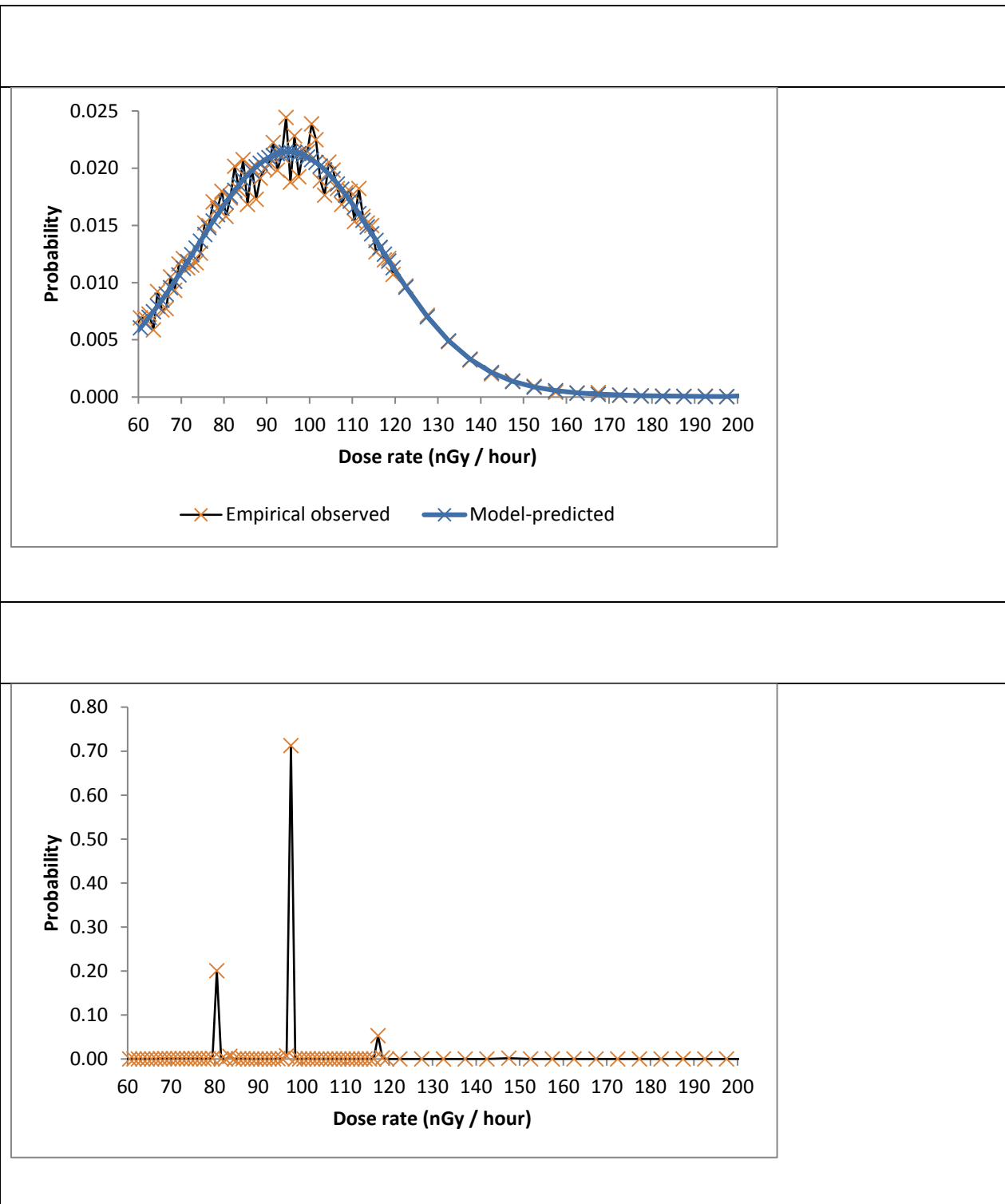
$$L[(p_i)_{i=1}^N, \sigma] = \sum_{j=1}^M n_j \ln[F(d_j | (p_i)_{i=1}^N, (\mu_i)_{i=1}^N, \sigma) - F(d_{j-1} | (p_i)_{i=1}^N, (\mu_i)_{i=1}^N, \sigma)] \quad (D2)$$

The results of maximizing (D2) with the p_i corresponding to dose-rates above 160 nGy/h constrained equal for stability, yields an estimate of the probabilities to be attached to each

component Normal distribution, as given by the p_i , which are plotted in Fig D1. As can be seen, the dose-rate distribution is largely given by a combination of three Normal distributions $X_i \sim N(\mu_i, \sigma^2)$ with means at 80.5, 97.5 and 117.5 nGy/h, with weights $p_i = 0.198, 0.706, 0.052$ respectively. The model provides a good fit to the observed dose-rate distribution. The common estimated SD was $\sigma=19.79$ nGy/h.

While the choice of fitting three Normal distributions was essentially arbitrary and for illustrative purposes, it is plausible that the two Normal distributions with lower means represent the underlying bimodality while the third reflects a slight shoulder on the high dose side of the distribution.

Fig D1 (a) Empirical and model-fitted distributions and (b) probabilities $(p_i)_{i=1}^N$ associated with component Normal distributions.



Appendix E: Distributions of indoor gamma-ray dose-rates in other countries

Population distributions of indoor gamma-ray dose-rates are discussed in Annex B of the 2000 Report of the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR, 2000b); Table 12 of that report presents data for eleven countries. UNSCEAR (UNSCEAR, 2000b) notes that “The distribution in Italy is also wide and approximately bimodal. The distributions in the Russian Federation, Finland and Lithuania are characterized by separate peaks in the distributions at decades 2 or 3 above the country mean.” (In the case of Lithuania, the two peaks are presumably too close together to be resolved in the data presented, or else one may lie in the substantial tail above 100 nGy/h.)

The indoor gamma-ray dose-rate distribution for Denmark, while apparently unimodal in the UNSCEAR (UNSCEAR, 2000b) tabulation, is bimodal when examined at a finer resolution (Ulbak *et al*, 1987; Ulbak *et al*, 1988). In fact, Ulbak and co-workers (Ulbak *et al*, 1987; Ulbak *et al*, 1988) made measurements in 489 buildings and analysed the data for one-family houses and apartment blocks separately.

The distribution of indoor gamma-ray dose-rates in Spain also exhibits bimodality, with distinct peaks at 60-69 and 80-89 nGy/h (44% of measurements were >100 nGy/h). This bimodality would provisionally be assigned to geological factors (Marta García-Talavera, private communication).

The dose-rate distribution in Hungary is also bimodal, with a low dose-rate peak at 20-29 nGy/h, then a steady rise to 90-99 nGy/h, and a second and much larger peak somewhat greater than 100 nGy/h. The data for Bulgaria exhibit distinct peaks at 60-69 nGy/h and 80-89 nGy/h.

Thus, at least seven countries (Denmark, Finland, Hungary, Italy Lithuania, the Russian Federation and Spain), of the eleven for which UNSCEAR (UNSCEAR, 2000b) presents distributions of indoor gamma-ray dose-rates, have bimodal distributions; it is not clear whether the other four do so. In particular, the data for Bulgaria, while rather sketchy, do hint at bimodality. The Belgian distribution, while unimodal, is based on only 100 measurements (Uyttenhove *et al*, 1984) and any structure may be difficult to see. The data for Romania appear to be based on theoretical calculations rather than direct measurements (Iacob & Botezatu,

2004), and while this may well give a good estimate of the mean population dose it is likely to overlook any finer structure that might be present.

Other reports of indoor gamma-ray dose-rates are worthy of note. In particular, Schmier and co-workers (Schmier *et al*, 1982) summarised a large study in West Germany in which 30,000 measurements of exposure were made in dwellings (generally three measurements per building). Differences were found between regions (“Länder”) and between buildings constructed with different materials. However, the overall distribution of exposure-rates was reported in fairly broad categories and it is likely that any deviations from unimodality would not have been apparent.

Mjönes (Mjönes, 1986) reported from a survey of 1,300 Swedish dwellings that dose-rates in apartments were twice those in single family houses. This no doubt accounts for the bimodality in the overall distribution, and there were also substantial regional differences. The Swedish dwellings with the highest indoor gamma-ray dose-rates tended to be those in which alum shale aerated concrete had been used as a building material (Mjönes, 1986) (Axelson *et al*, 2002).

Storruste and co-workers (Storruste *et al*, 1965), in a study of 2,026 Norwegian dwellings reported substantially lower indoor gamma-ray dose-rates in buildings made of wood as compared to concrete or brick. Note that alum shale aerated concrete was not used in Norway.

It is clear that bimodality may be present, but not apparent, if insufficient measurements have been made or if the tabulated data are too coarsely stratified. To this the situation which we report in Great Britain should be added where the underlying bimodality is obscured by inter-house variations and measurement errors.

The reasons for bimodality are likely to be characterised by specific features in each country. In the case of Italy, the broad bimodality is ascribed to two Regions in Central Italy (Lazio and Campania, out of a total of 21) having conspicuously high dose-rates (Bochicchio *et al*, 1996). It appears that geology and the distribution of population by geology drives the differences, at least in part because local materials are used for building (Bochicchio *et al*, 1996). In Spain the bimodality is believed to correspond to different geologies. In Denmark and Finland it appears to correspond to two types of dwelling, but in Finland there are also strong geographical variations which are ascribed to varying radionuclide concentrations in soils and construction materials (Vesterbacka, 2015). As noted in the main text, in Great Britain a number of factors seem to be operating, none of which is dominant.

In the general situation, indoor gamma-ray dose-rates may vary depending on local geology, building materials and house styles; there is thus no obvious reason why the indoor distributions should be bimodal rather than multimodal. The Hungarian distribution looks more complex than simply bimodal, and it may be that in other countries further complexity is concealed by limitations on the accuracy of measurements and the number of homes measured. It is also unclear why bimodality is shown by some original measurement datasets, but is only obvious in the UK data when measurements are averaged over comparatively small geographical areas. Of course, in some countries the separation of the peaks is greater than in the UK data, which makes it easier to distinguish the separate modes. It is also likely that the bimodality (or greater complexity) may be more marked in some countries than others.

Appendix F: The influence of building materials on indoor gamma-ray dose-rates

The indoor gamma-ray dose-rate in a building may be regarded as the outdoor dose-rate before the house was constructed, reduced by the shielding provided by the building, but increased again by emissions from radionuclides within the building materials. These emissions are normally the largest contribution to the indoor dose-rate as noted by the UNSCEAR 2000 Report, Paragraph 57, (UNSCEAR, 2000b), at least for buildings constructed of conventional masonry materials, as are a large majority of those in the UK (Department for Communities and Local Government, 2010). Gamma-ray emissions from building materials come largely from K-40 and radionuclides in the U-238 and Th-232 decay chains, the three contributions being roughly equal, UNSCEAR 2000 Report, Annex B, Paragraph 43 (UNSCEAR, 2000b). These radionuclides are ubiquitous in the environment, but concentrations vary from one material to another. The extent to which building materials with high radionuclide concentrations might lead to high indoor gamma-ray dose-rates has been reviewed (European Commission, 1997) (European Commission, 1999).

Building materials apart from timber are derived from “earth materials” such as clays or rocks. In some cases, building materials are made from by-products of other processes, for example the use of power station coal ash in building blocks. The radioactivity in building materials will be determined by the radioactivity in the materials from which they are made, although the activity may be modified in the fabrication process. There will also be variation in the

radioactive content of different material from the same broad geological strata. Relatively few data have been published on radioactivity in British building materials. A review by the European Commission (European Commission, 1997) cites three publications (Cliff *et al*, 1984; Hamilton, 1971; O'Riordan & Hunt, 1977) and an unpublished contract report. Models are available to calculate indoor gamma-ray dose-rates given the activity concentrations in the building materials (Markkanen, 1995) (Risica *et al*, 2001) (de Jong & van Dijk, 2008). However, the lack of information on the materials used in the construction of the dwellings of interest precludes their application here.

A detailed analysis has been published of the radiological consequences of using building materials (blocks, but possibly also concrete) made using coal ash (bottom or fly). Smith and co-workers (Smith *et al*, 2001) state that power station ash amounts to about 16% by weight of the coal that is burned and that there is a considerable concentration of radionuclides in it. Smith and co-workers go on to estimate (section 5.1) that “the total external dose to a resident of a building constructed using building materials containing ash, from all radionuclides in the material, is approximately 893 $\mu\text{Sv.y}^{-1}$. The corresponding dose in a building constructed from similar materials that do not contain ash is approximately 758 $\mu\text{Sv.y}^{-1}$.”

The increased dose is thus $\sim 135 \mu\text{Sv.y}^{-1}$, approaching 20% of the total dose. However, there does not appear to be published information on the proportion of buildings constructed using coal ash-based materials, nor on where these materials were used. It is plausible that such building materials are becoming increasingly common with pressures to reduce the amount of material sent to land-fill. However, it is likely that coal might be transported considerable distances from mine to power station (and these coal ash blocks are a relatively recent innovation) so that these building materials are not used particularly close to where the coal was mined. It is interesting to note that the use of coal ash in building materials is not an exclusively modern practice; in the London area ‘town ash’ (residue from domestic coal-burning) was used for brick making in the 19th and 20th century (Bloodworth, 2016).

Bricks, perhaps the most common building material in Great Britain, were generally made close to the source of raw materials (clay or shale), though this became a less powerful tendency over the twentieth century as production became more centralised (Brunskill, 1997). Brick works also tended to be positioned close to sources of coal to provide fuel. In the early days bricks were used close to where they were made (indeed, initially, bricks were made at the

construction site). With the introduction of canals, railways and road transport bricks were used further from the source of the raw material (Brunskill, 1997).

In Great Britain the main brick-making raw materials have been derived from material from (DCLG (Department for Communities and Local Government), 2007) (Scottish Executive, 2007) (Bloodworth, 2016) Carboniferous Coal Measures, Triassic Mercia Mudstones, Jurassic Oxford clay; and Cretaceous clay. The first two of these are older than the second pair. Dwellings built on these older strata have, on average, higher mean indoor gamma-ray dose-rates (104 and 101 vs 85 and 85 nGy/h see Table 1). There is, however, considerable variation within each bedrock class.

If houses tended until relatively recently to be built using bricks from the nearest source of clay, then dwellings in the English Midlands and Northwest England conurbations will tend to have indoor gamma-ray dose-rates similar to those in houses built on the Carboniferous Coal Measures and Triassic Mercia Mudstones, whilst dwellings in the Northeast conurbations of England and the Central Valley of Scotland will tend to have indoor gamma-ray dose-rates similar to those in houses built on the Carboniferous Coal Measures. Conversely, indoor gamma-ray dose-rates in dwellings in London and the South-east of England are likely to be more similar to those built on the Jurassic Oxford clay and the Cretaceous clays. This provides a possible explanation for at least part of the general geographical variation in indoor gamma-ray dose-rates seen in OL-2b and OLR-2c. However, information on the sources and radioactive content of British building materials is very limited and firm deductions are not possible.

Appendix G: Other approaches to estimating indoor gamma-ray dose-rates in unmeasured homes

Introduction

In this paper and its predecessors (Chernyavskiy *et al*, 2016; Kendall *et al*, 2016b) methods for estimating indoor gamma-ray dose-rates in unmeasured dwellings have been developed and tested. A main aim of the present work was to predict dose quantities for use in epidemiological studies of the effect of natural background gamma radiation, but the work has scientific interest in its own right. The data available for developing models consisted of a set of 10,199

measurements in dwellings. The present task was greatly simplified by the fact that the individual measurements of the quantity in question were available (rather than just a summary) and by the large size of the measurement set. This resulted in a closely spaced grid of measurements such that most points at which predictions were required were close to a measurement location. The mean separation of measurement locations from their nearest neighbour was 1.2 km and the mean separation of the birth location of study subjects from the nearest measurement location was 1.1 km. However, these means were inflated by occasional large separations. Table G1 shows that well over a third of locations had a measurement within 500 m, about 70% within 1000 m and over 85% within 2,000 m

In many countries, including most of those considered here, the majority of people spend most of their time indoors (UNSCEAR, 2000a). In theory a more accurate estimate of doses from gamma-rays would be obtained by combining estimates of indoor and outdoor dose-rates with the respective occupancy factors. However, a reasonable approximation (e.g. (Kendall *et al*, 2006)) is that the fraction of time spent outdoors is small, the dose-rates indoors and outdoors are not greatly dissimilar, and that the mean overall dose-rate will not differ greatly from the indoor rate.

Another point to consider is whether it matters that the measurements on which the dose prediction method is based were taken during the period of epidemiological follow-up (in the present case, 1962-2010). It is true that indoor gamma-ray dose-rates are not completely constant in time. Thus, Minato (Minato, 1980) reported that atmospheric radon daughter concentrations and rainfall play an important role in short- and middle-term variations in the background radiation flux, while changes in soil dryness contribute to longer-term variations. However, such changes are largely temporary, and it is argued here that mean indoor gamma-ray dose-rates in existing homes do not change much with time over the timescale of interest for epidemiology. It is true that dose-rates in an older house may change if, say, it is replastered with gypsum containing relatively high concentrations of uranium. But generally constant concentrations of radionuclides in surrounding geology and in the building materials mean that gamma radiation dose-rates in a house will be roughly constant, much as they are outdoors. This is not necessarily the case with radon, for example, where changes in heating and in ventilation can affect in-door radon concentrations significantly.

As described in the main paper, in Great Britain (GB) the best predictive method was found to be a linear combination of various simple models - averages over small areas, over geological

units or a weighted sum of the nearest measurements. Geology was not a very powerful predictor, though it may be the best single predictor available. Despite testing several detailed models, with the available data, substantial Mean Square Error was associated with the optimal fit. Geostatistical (“kriging”) methods performed slightly less well; possible reasons for this are explored in the main text.

Other approaches

A French study of natural background radiation and childhood cancer (Demoury *et al*, 2017) made use of dose estimates by Warnery *et al* (Warnery *et al*, 2015). Like the GB study, Warnery and co-workers also had a large set of indoor measurements made in 17,404 dental surgeries and veterinary clinics. Two kinds of variogram-based geostatistical modelling were conducted to estimate indoor terrestrial gamma-ray dose-rates in France on a grid of 1 km squares: ordinary kriging (considering only the locations of the measurement points); and multi-located cokriging, (also considering the geogenic uranium potential of the measurement locations). Warnery and co-workers used cross-validation in which single observed measurements were successively excluded from the total measurement set and then predicted using the remaining 17,403 measurements (Warnery *et al*, 2015) (Marquant *et al*, 2018). The Mean Square Errors of the two methods calculated in this way were 409 and 407 (nSv/h)² respectively. The arithmetic mean was 76 nSv/h.

In spatial statistics the total variance can be described as the sum of the “sill” and the “nugget effect”. The former is the variation that is explained by distance between observations; the latter is random variation that is non-spatial. Outside the context of spatial statistics a nugget effect is just “random error”. The variogram-based modelling approach of Warnery *et al* yields a nugget effect of 35% of the total variance of the 17,404 available measurements. Warnery and co-workers (Warnery *et al*, 2015) considered it likely that the nugget effect could mainly be due to the influence of local factors that are not taken into account by the modelling (i.e. inter-house variation) rather than metrological inaccuracies.

Kendall *et al* (2018) queried whether the distribution of dose-rates in French dental surgeries and veterinary practices was necessarily similar to that in French homes (Kendall *et al*, 2018). However, whatever the case may be, it is unlikely to affect conclusions drawn from the experience of Warnery and co-workers (Warnery *et al*, 2015) in fitting the data.

Two other European epidemiological studies, in Finland (Nikkilä *et al*, 2016) and in Switzerland (Spycher *et al*, 2015), were less well-placed than France and GB in that they had available only pre-existing maps of outdoor dose-rates.

The Finnish outdoor gamma radiation results were based on measurements from vehicles covering 15,000 km during 1978–1980 (Akima, 1978; Arvela *et al*, 1995). The co-ordinates of the midpoints of 410 representative, evenly distributed, sections were used as the co-ordinates for the measurements. The SAS G3GRID procedure was used for interpolating values from an irregularly spaced set of points, generating a rectangular grid of 8 x 8 km². G3GRID uses a modification of the bivariate triangular interpolation method of Akima (Akima, 1978). However, the Finnish workers also had available measurements of indoor gamma-ray dose-rates in 346 dwellings and also access to a housing register which allowed the homes of study subjects to be classified as a house or flat. Dwelling-type specific conversion factors were then used to estimate indoor gamma-ray dose-rates from the outdoor values.

Spycher *et al* (2015) made use of the Swiss outdoor natural terrestrial radiation survey (Rybach *et al*, 2002) which combined a variety of measurements: airborne gamma-ray spectrometry (about 10% of the country's surface surveyed by helicopter); in situ gamma-ray spectrometry (166 sites); in situ dose-rate measurements using ionization chambers (837 sites); and laboratory measurements of rock and soil samples from 612 sites. These measurements were made from the early 1960s to mid-1990s. In addition to airborne measurements, a total of 1,615 ground data points were available, which corresponds to about one point per 25 km². Dose-rates for cells of a 2 × 2 km grid were interpolated from the available data points using the inverse distance method and a search radius of 12 km. As in most countries, no housing register was available, and the Swiss researchers analysed their epidemiological data in terms of these outdoor dose-rate estimates.

The three studies described so far had to make use of such pre-existing data as were available. However, two epidemiological studies comparing areas of high natural background with control areas were able to undertake special surveys and measurements to assess exposures. These studies were set in Guangdong Province, China (Tao *et al*, 2012), and in Kerala, India (Nair *et al*, 2009). In both cases extensive sets of indoor and outdoor gamma-ray dose-rate measurements were made, and habit surveys gave sex- and age-specific house occupancy factors. The researchers were thus able to make estimates of doses based on near-by measurements. These were not necessarily based on measurements for the individual

concerned, but sometimes on village-specific parameters. Nevertheless, they were based on direct measurements of the quantities in question at locations close to the homes of study subjects. Moreover, measurements using personal dosimeters worn by a sample of local inhabitants were undertaken and used to validate the estimated doses. Agreement was reported to be good, but in China comparisons appear to have been of group averages rather than individuals.

Summary

Different workers have used very different approaches, driven by the data and resources available. In the Introduction to this paper the E-OLS model favoured in the current paper is contrasted with the likelihood-based estimation as in the MRGP models favoured by Chernyavskiy and co-workers (Chernyavskiy *et al*, 2016) and the variogram-based (co-)kriging models by Warnery and co-workers (Warnery *et al*, 2015). It is likely that developments in methodology and in computer hardware will increase the role for geostatistical models in the future (Heaton *et al*, 2017).

Both the French and the British investigations found that even their optimal models left large non-spatial inter-house variations that could not be explained with the explanatory variables available. However, approaches like those of the French and British investigators are dependent on access to sets of individual measurements. Other methods must be adopted where such data are not available. Conversely, if it is possible to undertake a significant number of new measurements then a much more detailed approach is possible.

Table G1 Separations from nearest measurement location of measurement locations (excluding that in question) and birthplaces of study subjects

Distance band (m)	Measurement Locations		Birth Places	
	Cumulative Frequency	%	Cumulative Frequency	%
<500	3,751	36.8	4,880	39.2
<1000	6,951	68.2	8,865	71.2
<1500	8,248	80.9	10,384	83.4
<2000	8,746	85.8	10,985	88.3
<2500	9,049	88.7	11,308	90.9
<3000	9,252	90.7	11,526	92.6
<5000	9,750	95.6	12,019	96.6

<10000	10,109	99.1	12,368	99.4
Totals	10,199		12,446	

Figure Captions

Fig 1 Distribution of indoor gamma-ray dose-rate measurement data, nGy/h

Fig 2 Distributions of the numbers of measurements ($n=10,199$) in county districts (CDs) and in 10 km grid squares by mean indoor gamma-ray dose-rate for that areal unit (i.e. where measurements are represented by the mean for the CD or for the 10 km grid square)

Fig 3 Geographical distribution of indoor gamma-ray dose-rate measurements below 80 and above 110 nGy/h

Fig 4 Distributions of indoor gamma-ray dose rate measurements (nGy/h) averaged over county districts (CDs), and predictions from E-OLS, Gaussian-Matérn (GM), MRGP-8-3 and MRGP-8-5 models at measurement locations

Table Captions

Table 1 Indoor gamma-ray dose-rates (nGy/h) for Bedrock Classes and Reduced Classes with minimum, maximum and Quartile values

Table 2 Summary statistics for fitting various models to the measured indoor gamma-ray dose-rates (nGy/h); CD – county districts

Table 3 Analysis of variance for optimal E-OLS model

References

- Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Petrov BN, Czaki F (eds), pp 267-281. Budapest: Akadémiai Kiadó
- Akaike H (1981) Likelihood of a model and information criteria. *J Econometrics* 16(1): 3-14
- Akima H (1978) A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points. *ACM Trans Math Softw* 4(2): 148-159
- Appleton JD, Cave MR (2018) Variation in soil chemistry related to different classes and eras of urbanisation in the London area. *Appl Geochem* 90: 13-24
- Arvela H, Hyvönen H, Lemmelä H (1995) Indoor and outdoor gamma radiation in Finland. *Radiat Protect Dosim* 59: 25-32
- Axelsson O, Fredrikson M, Åkerblom G, Hardell L (2002) Leukemia in childhood and adolescence and exposure to ionizing radiation in homes built from uranium-containing alum shale concrete. *Epidemiology (Cambridge, Mass)* 13(2): 146-150
- Bloodworth A (2016) *The influence of geology on English brickmaking* London: Institute of Historical Building Conservation.
- Bohicchio F, Campos Venuti G, Monteventi F, Nuccetelli C, Piermattei S, Risica S, Tommasino L, Torri G (1996) Indoor exposure to gamma radiation in Italy. *Proceedings of the 9th International Congress of the IRPA, Vienna, April 1996* 2: 190-192
- Brunskill RW (1997) *Brick building in Britain* London: Victor Gollancz in association with Peter Crawley
- Chernyavskiy P, Kendall GM, Wakeford R, Little MP (2016) Spatial prediction of naturally occurring gamma radiation in Great Britain. *J Environ Radioact* 164: 300-311
- Cliff KD, Miles JCH, Brown K (1984) *NRPB R-159 The incidence and origin of radon and its decay products in buildings* Chilton.
- Davison AC, Hinkley DV (1997) *Bootstrap methods and their application*. Cambridge: Cambridge University Press
- DCLG (Department for Communities and Local Government) (2007) *Brick Clay: Mineral Planning Factsheet*, Government DfCal (ed). London: Department for Communities and Local Government
- de Jong P, van Dijk JW (2008) Calculation of the indoor gamma dose rate distribution due to building materials in the Netherlands. *Radiat Prot Dosim* 132(4): 381-9
- Demoury C, Marquant F, Ielsch G, Goujon S, Debayle C, Faure L, Coste A, Laurent O, Guillevic J, Laurier D, Hémon D, Clavel J (2017) Residential Exposure to Natural Background Radiation and Risk of Childhood Acute Leukemia in France, 1990-2009. *Environmental Health Persp* 125(4): 714-720
- Department for Communities and Local Government (2010) *English Housing Survey Housing stock report 2008*.
- Diggle PJ, Ribeiro PJ, Jr (2007) Model-based geostatistics. In *Springer Series in Statistics* pp 1-228. 233 Spring Street, New York, NY 10013: Springer

Driscoll CMH, Green BMR, McKinlay AF, Richards DJ (1983) Environmental Radiation Measurements using Lithium Fluoride Thermoluminescent Dosemeters. *Radiat Prot Dosim* 6(1-4): 241-244

Dudley Stamp L (1931) Land Utililisation Survey of Britain. *The Geographical Journal* 78: 40-47

Efron B (1987) Better bootstrap confidence intervals. *J Am Statist Assoc* 82(397): 171-185

Environment Agency (2010) *GIS Classification of the 1930-1940's Land Utilisation Survey Entec Final Report June 2010*.

European Commission (1997) *Radiation protection 96 Enhanced radioactivity of building materials* Luxembourg: Office for Official Publications of the European Communities.

European Commission (1999) *Radiation protection 112 Radiological Protection Principles concerning the Natural Radioactivity of Building Materials* Luxembourg.

Green BMR, Lomas PR, Bradley EJ, Wrixon AD (1989) *Gamma-radiation Levels Outdoors in Great Britain. NRPB - R191*. Didcot: National Radiological Protection Board

Hamilton EI (1971) The Relative Radioactivity of Building Materials. *Am Indus Hygiene Assoc J* 32(6): 398-403

Heaton M, Datta A, Finley A, Furrer R, Guhaniyogi R, Gerber F, Hammerling D, Katzfuss M, Lindgren F, Nychka D, Zammit-Mangion A (2017) Methods for Analyzing Large Spatial Data: A Review and Comparison *arXiv* 1710.05013

Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. *Biometrika* 76(2): 297-307

Iacob O, Botezatu E (2004) Population Exposure to Natural Radiation Sources in Romania. *Proceedings of the 11th IRPA* 6a33

Kendall GM, Hughes JS, Oatway WB, Jones AL (2006) Variations in radiation exposures of adults and children in the UK. *J Radiol Prot* 26: 257-276

Kendall GM, Little MP, Wakeford R (2018) Comment on “Indoor terrestrial gamma dose rate mapping in France: A case study using two different geostatistical models” by Warnery et al. (*J. Environ. Radioact.* 2015, 139, 140-148). *J Environ Radioact* 182: 172-173

Kendall GM, Little MP, Wakeford R, Bunch KJ, Miles JCH, Vincent TJ, Meara JR, Murphy MF (2013) A record-based case-control study of natural background radiation and the incidence of childhood leukaemia and other cancers in Great Britain during 1980-2006. *Leukemia* 27(1): 3-9

Kendall GM, Miles JCH, Rees D, Wakeford R, Bunch KJ, Vincent TJ, Little MP (2016a) Variation with socioeconomic status of indoor radon levels in Great Britain: The less affluent have less radon. *J Environ Radioact* 164: 84-90

Kendall GM, Wakeford R, Athanson M, Vincent TJ, Carter EJ, McColl NP, Little MP (2016b) Levels of naturally occurring gamma radiation measured in British homes and their prediction in particular residences. *Radiat Environ Biophys* 55(1): 103-124

- Kendall GM, Wakeford R, Bunch KJ, Vincent TJ, Little MP (2015) Residential mobility and associated factors in relation to the assessment of exposure to naturally occurring radiation in studies of childhood cancer. *J Radiol Prot* 35: 835-868
- Markkanen M (1995) *Radiation Dose Assessments for Materials With Elevated Natural Radioactivity Report STUK-B-STO 32* Helsinki: STUK Finnish Centre for Radiation and Nuclear Safety.
- Marquant F, Demoury C, Ielsch G, Laurier D, Hémon D, Clavel J (2018) Response to comment on "Indoor terrestrial gamma dose rate mapping in France: A case study using two different geostatistical models" by Warnery et al. *J Environ Radioact* 182: 174-176
- Matérn B (1960) Spatial variation. In *Meddelanden fran Statens Skogsforsknings institut*, Vol. 49, pp 1-144. Stockholm:
- Minato S (1980) Analysis of Time Variations in Natural Background Gamma Radiation Flux Density. *J Nuc Sci Tech* 17(6): 461-469
- Mjönes L (1986) Gamma Radiation in Swedish Dwellings. *Radiat Prot Dosim* 15(2): 131-140
- Nair RR, Rajan B, Akiba S, Jayalekshmi P, Nair MK, Gangadharan P, Koga T, Morishima H, Nakamura S, Sugahara T (2009) Background radiation and cancer incidence in Kerala, India-Karanagappally cohort study. *Health Phys* 96(1): 55-66
- Nikkilä A, Erme S, Arvela H, Holmgren O, Raitanen J, Lohi O, Auvinen A (2016) Background radiation and childhood leukemia: A nationwide register-based case-control study. *Int J Cancer* 139(9): 1975-82
- Nychka D (2014) *R package LatticeKrig: multiresolution kriging based on Markov random fields. Version 3.4. Comprehensive R Archive Network (CRAN)*. R project.
- Nychka D, Bandyopadhyay S, Hammerling D, Lindgren F, Sain S (2015) A Multiresolution Gaussian Process Model for the Analysis of Large Spatial Datasets. *J ComputGraph Stat* 24(2): 579-599
- O'Riordan MC, Hunt GJ (1977) Radiological controls for construction materials. *Proc 4th IRPA, Paris* 3: 1053-1056
- Pearson K (1895) VII. Note on regression and inheritance in the case of two parents. *Proc Roy Soc London* 58(347-352): 240-242
- R Core Team (2017) R Project version 3.4.3 (2017): A Language and Environment for Statistical Computing Vol. 2017. Vienna, Austria: R Foundation for Statistical Computing
- R Project version 3.2.2 (2015) Comprehensive R Archive Network (CRAN)
- R Project version 3.4.4 (2018) R: A language and environment for statistical computing. version 3.4.4 <https://www.r-project.org>. Vienna, Austria: R Foundation for Statistical Computing
- Rao CR (2002) *Linear statistical inference and its applications (2nd ed)*. New York: Wiley
- Risica S, Bolzan C, Nuccetelli C (2001) Radioactivity in building materials: room model analysis and experimental methods. *Science Tot Environ* 272(1-3): 119-126

- Rybach L, Bachler D, Bucher B, Schwartz GF (2002) Radiation doses of Swiss population from external sources. *J Environ Radioact* **62**: 277-286
- Schmier H, Kolb W, Mehl J, Oberhausen E (1982) The Variation of Natural Background Radiation in the Federal Republic of Germany. In *Second Special Symposium on the Natural Radiation Environment 1981*, Vohra KG, Mishra UC, Pillai KC, Sadasivan S (eds) pp 173-179. Bombay, India: John Wiley and Sons
- Scottish Executive (2007) Brick clay Geology and mineral planning factsheets for Scotland, Executive S (ed). Edinburgh: Scottish Executive
- Shaw KB, Wall BF (1977) *NRPB-R65. Performance tests on the NRPB Thermoluminescent Dosemeter*. Chilton: NRPB. Report no. NRPB-R65
- Smith KR, Crockett GM, Oatway WB, Harvey MP, Penfold JSS, Mobbs SF (2001) *Radiological Impact on the UK Population of Industries Which Use or Produce Materials Containing Enhanced Levels of Naturally Occurring Radionuclides: Part I: Coal-fired Electricity Generation* Chilton: National Radiological Protection Board.
- Southall H, Aucott P, Baily B (2007) *1930s Land utilisation mapping: An improved evidence-base for policy? Environment Agency Science Report: SC050031* London: Environment Agency.
- Spearman C (1904) The proof and measurement of association between two things. *Am J Psychology* **15**(1): 72-101
- Spycher BD, Lupatsch JE, Zwahlen M, Roosli M, Niggli F, Grotzer MA, Rischewski J, Egger M, Kuehni CE (2015) Background Ionizing Radiation and the Risk of Childhood Cancer: A Census-Based Nationwide Cohort Study. *Environ Health Perspect* **123** 622-628
- Storruste A, Reistad A, Rudjord T, Dahler A, Liestol I (1965) Measurement of Environmental Gamma Radiation in Norwegian Houses. *Health Phys* **11**: 261-269
- Tao Z, Akiba S, Zha Y, Sun Q, Zou J, Li J, Liu Y, Yuan Y, Tokonami S, Morishoma H, Koga T, Nakamura S, Sugahara T, Wei L (2012) Cancer and non-cancer mortality among Inhabitants in the High Background Radiation Area of Yangjiang, China (1979-1998). *Health Phys* **102**(2): 173-181
- UK Childhood Cancer Study Investigators (2000) The United Kingdom childhood cancer study: objectives, materials and methods. *Br J Cancer* **82**(5): 1073-1102
- UK Childhood Cancer Study Investigators (2002) The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionising radiation: 2: gamma radiation. *Br J Cancer* **86**(11): 1727-1731
- Ulbak K, Stenum B, Klinder O, Sørensen A, B. Majborn B, Bøtter-Jensen L, Nielsen SP (1987) *Naturlig straling i danske boliger* Copenhagen: National Board of Health
- Ulbak K, Stenum B, Sørensen A, B. Majborn B, Bøtter-Jensen L, Nielsen SP (1988) Results from the Danish Indoor Radiation Survey. *Radiat Prot Dosim* **24** (1-4): 401-405
- UNSCEAR (2000a) *Sources and effects of ionising radiation. United Nations Scientific Committee on the Effects of Atomic Radiation. 2000 Report to the General Assembly with Scientific Annexes* New York: United Nations.
- UNSCEAR (2000b) *United Nations Scientific Committee on the Effects of Atomic Radiation Sources and effects of ionising radiation. 2000 Report to the General Assembly with Scientific Annexes* New York: United Nations.

UNSCEAR (2010) *United Nations Scientific Committee on the Effects of Atomic Radiation 2008 Report. Annex B. Exposures of the public and workers from various sources of radiation* New York.

Uyttenhove J, Maryns R, Janssens A, Vanmarcke H, Jacobs R (1984) Survey on Natural Radiation in Houses in Belgium. *Radiat Prot Dosim* 7 (1-4): 275-278

Vesterbacka K (2015) *Surveillance of Environmental Radiation in Finland Annual Report 2014* Helsinki: STUK.

Warnery E, Ielsch G, Lajaunie C, Cale E, Wackernagel H, Debayle C, Guillevic J (2015) Indoor terrestrial gamma dose rate mapping in France: a case study using two different geostatistical models. *J Environ Radioact* 139: 140-8

Wendland H (1995) Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv Computat Math* 4(1): 389-396

Wrixon AD, Green BMR, Lomas PR, Miles JCH, Cliff KD, Francis EA, Driscoll CMH, James AC, O'Riordan MC (1988) *NRPB-R190. Natural radiation exposure in UK dwellings* Chilton, Didcot, Oxon: National Radiological Protection Board.