Radon, the lognormal distribution and deviation from it

Zornitza Daraktchieva and Jon Miles

Health Protection Agency, UK

10th International Workshop on the Geological Aspects of Radon Risk Mapping
22 – 25 September, 2010
Prague, Czech Republic

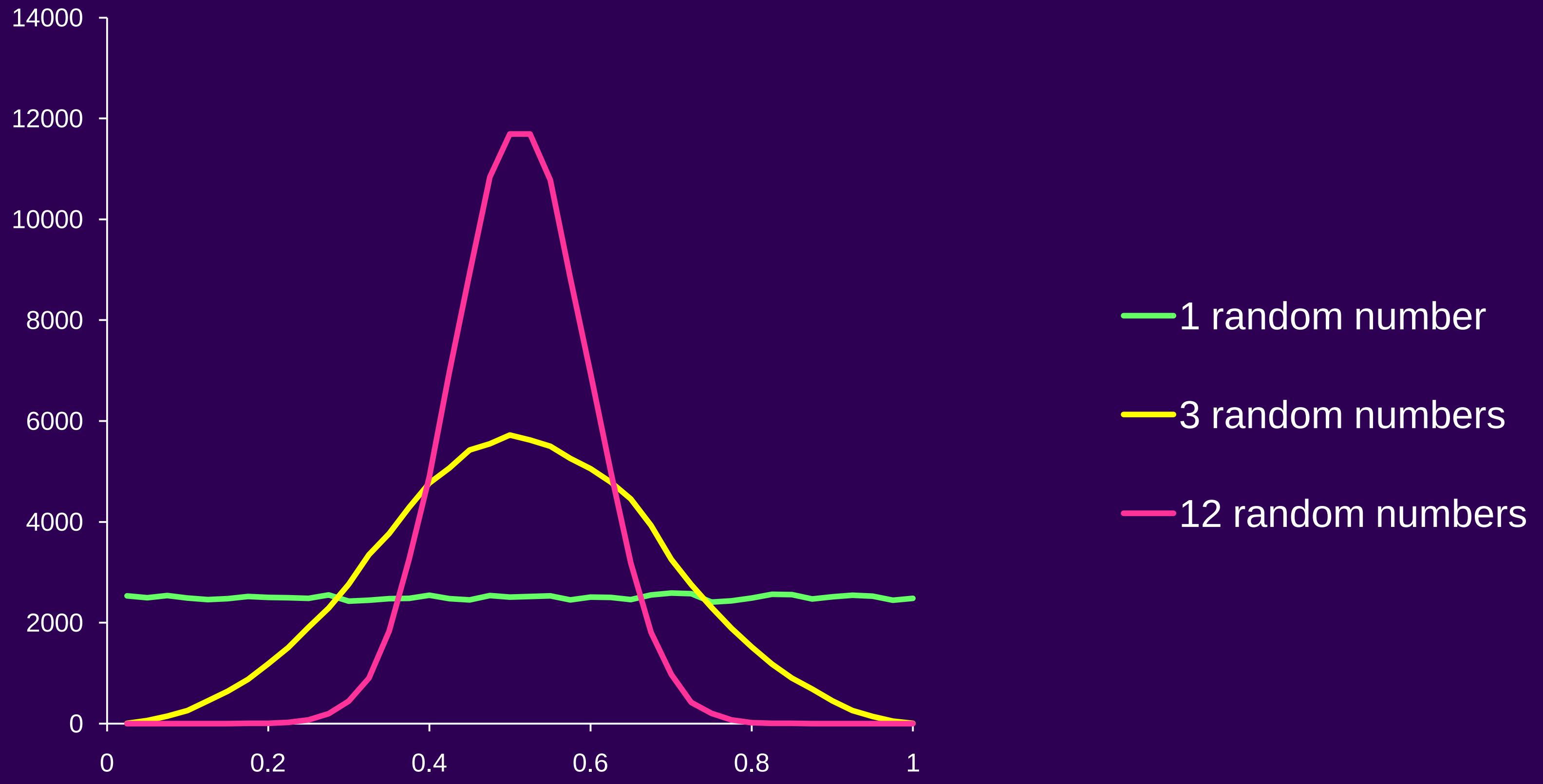
Why lognormal?

Many surveys of radon in homes in different countries have found the results to follow a lognormal distribution, or close to lognormal.

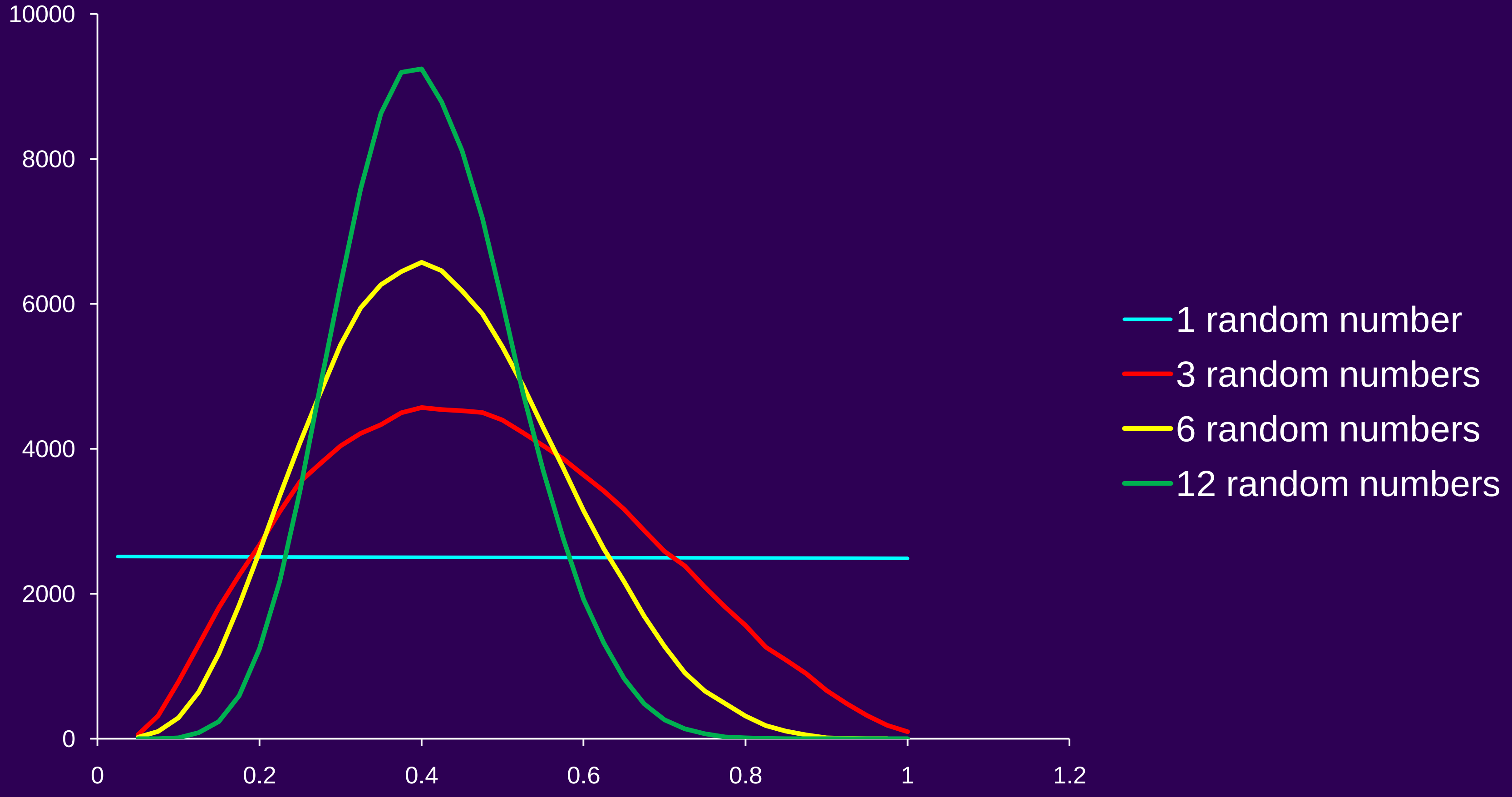
The reason for this can be understood in terms of the normal distribution:

- Values calculated by *summing* 12 or more random numbers follow a normal distribution
- Values calculated by *multiplying* 12 or more random numbers follow a *lognormal* distribution

Normal distribution from summing random numbers



Lognormal distribution from multiplying random numbers



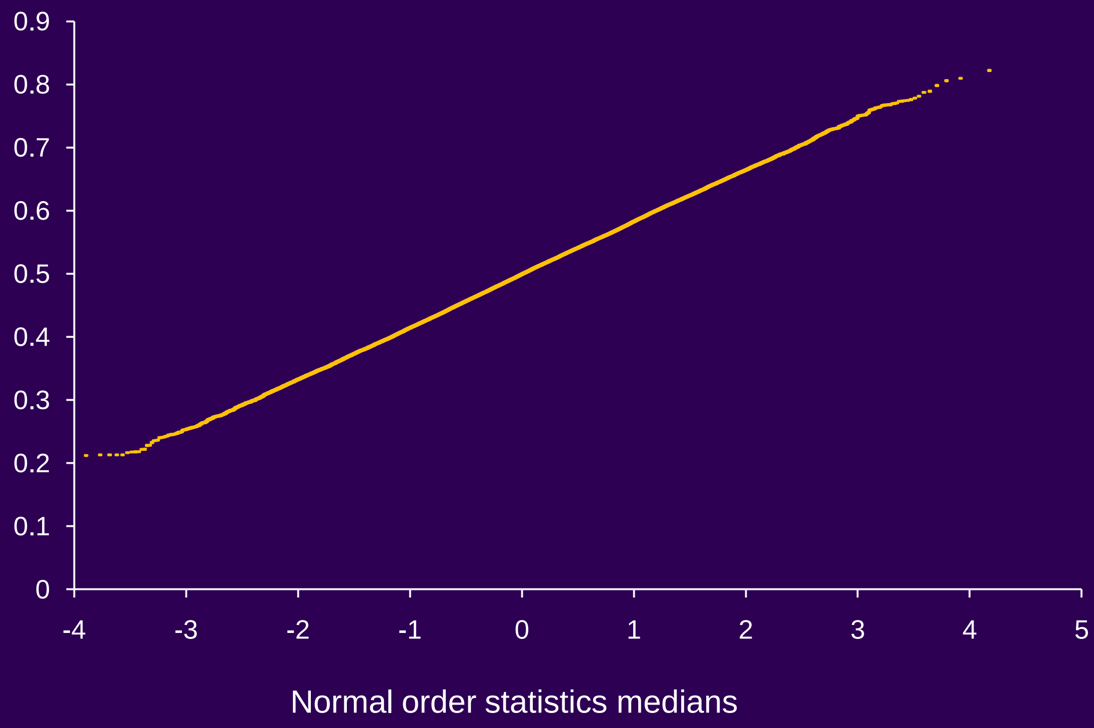
Normal probability plot

We use a graphical technique called the normal probability plot (or normal Quantile-Quantile, Q-Q plot) to assess whether the data set is normally distributed.

The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality.

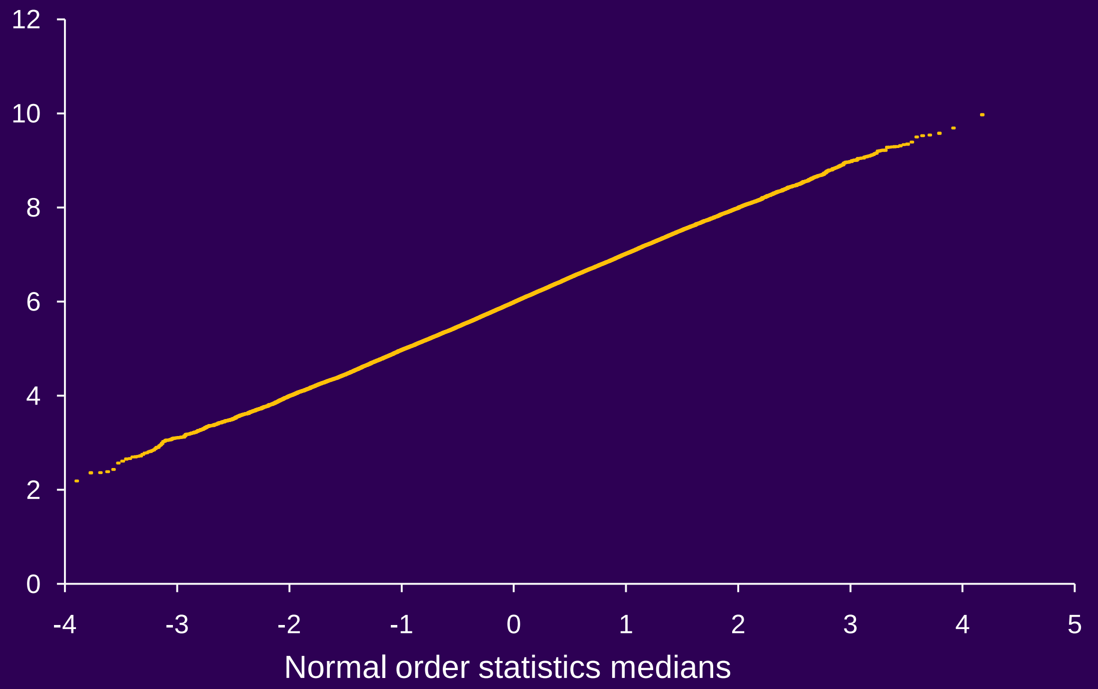
The line fitted to the data points gives the theoretical normal distribution with parameters corresponding to the mean (intercept) and standard deviation (slope).

Q-Q plot of normal distribution



This is the Q-Q plot of 32,000 points, each generated by summing 12 random numbers.

Variable magnitudes of random numbers



The random numbers do not have to be of the same magnitude – this is a plot of values where the random numbers varied over more than an order of magnitude

Radon from the ground into homes

Radon source: radium-226 in rocks and soils

Factors controlling how much radon enters a building:

- Ra-226 source concentration
- Grain size of source
- Moisture level around grains
- Permeability
- Entry routes into homes
- Underpressure in homes
- Ventilation rates
- Etc, etc . . . .

All are multiplying factors on the process of radon travel from origin to air in homes – hence expect lognormal distribution

But why should national distributions be lognormal?

The multiplying factors controlling radon entry can explain why radon concentrations in homes in one uniform area should be lognormally distributed.

But regional or national distributions must be sums of many local distributions.

So how do the local lognormal distributions sum to a different lognormal nationally?

The sum of many lognormal distributions may also be lognormal (but is not always so)

How many samples are there in the radon distribution?

We don’t know how many individual distributions might coexist in the measured radon distribution.

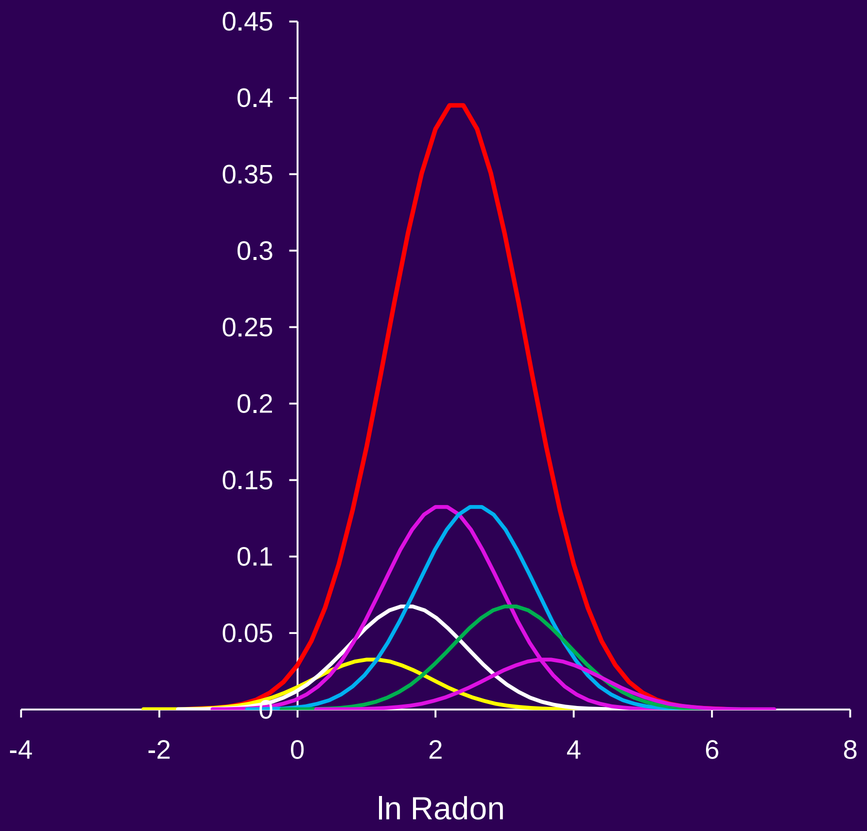
The measured radon can be expressed by multiplicative chain of various factors (J. Miles, Radiation Protection Dosimetry, 1994):

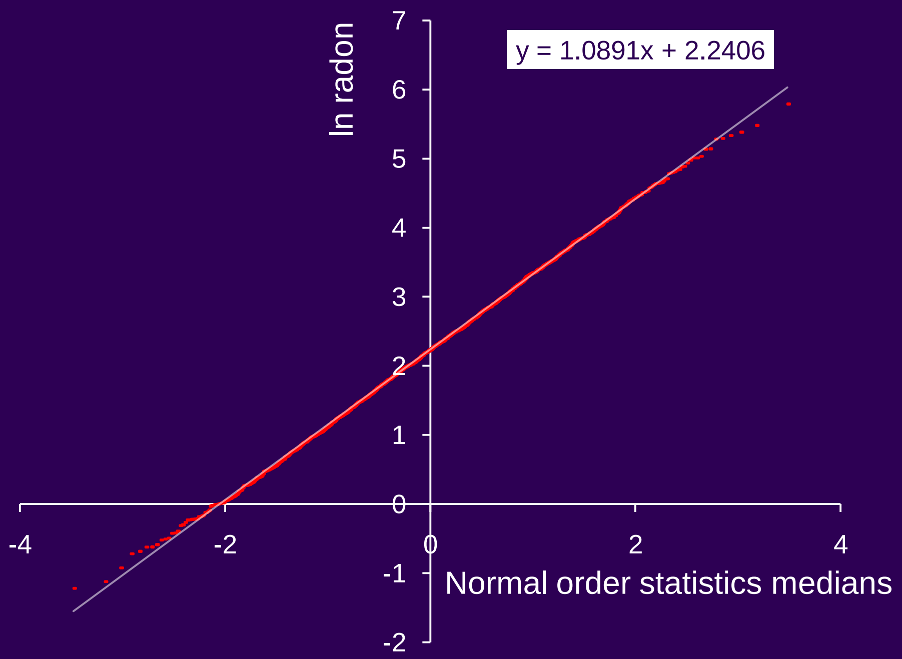
$$Ln(R_{net}) = ln(R_{soil}) + ln(A) + ln(B) + ln(C) + ln(D) + \ldots\ldots$$

We can show that the radon distribution can be presented as a linear combination of many different lognormal distributions.

Monte Carlo simulations of radon distribution 6 samples model

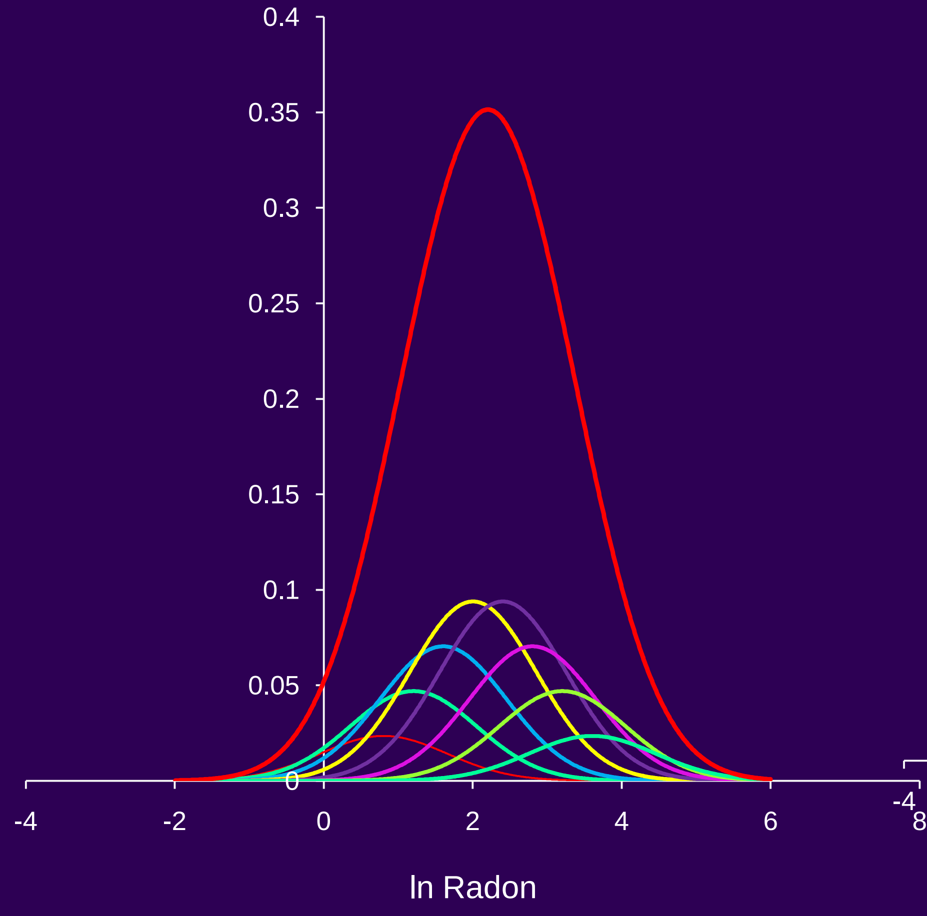
We have simulated 6 different weight samples with standard deviations of 0.85 and means of 1.0, 1.5, 2.0, 2.5, 3.0 and 3.5. The combined sample is a normal distribution with a mean of 2.2 and a standard deviation of 1.09 (red).

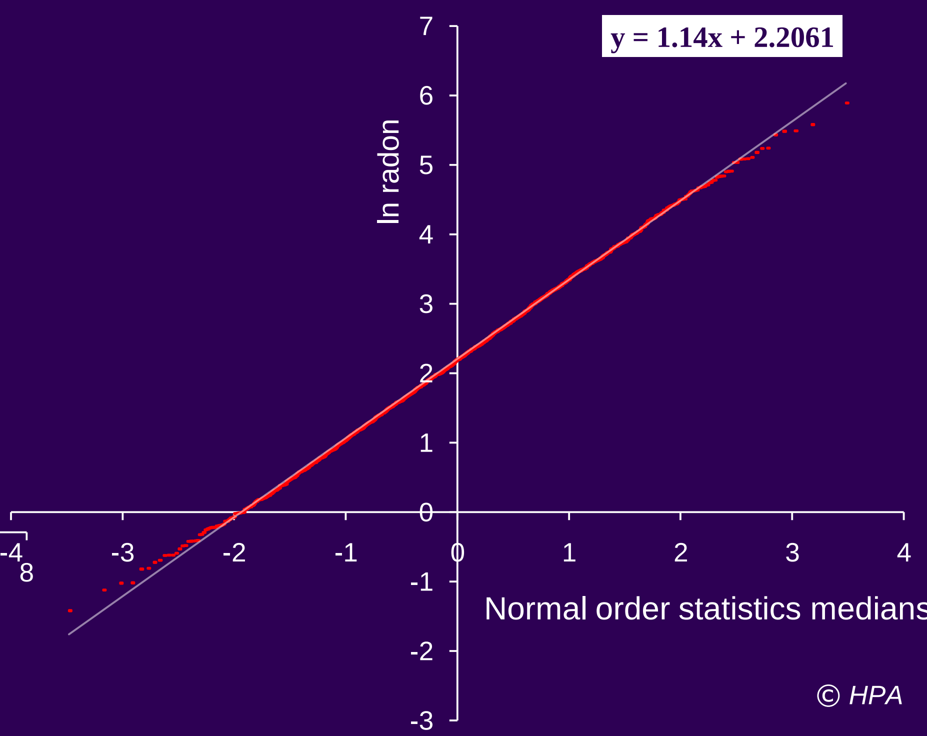




Monte Carlo simulations of radon distribution
8 samples model

8 different weight samples with standard deviations of 0.85 and means of 0.8, 1.2, 1.6, 2.0, 2.4, 2.8, 3.2 and 3.6. The result is a normal distribution with a mean of 2.2 and a standard deviation of 1.1 (red).





Comparison of lognormal model with UK radon distribution

National survey of the radon concentrations in UK

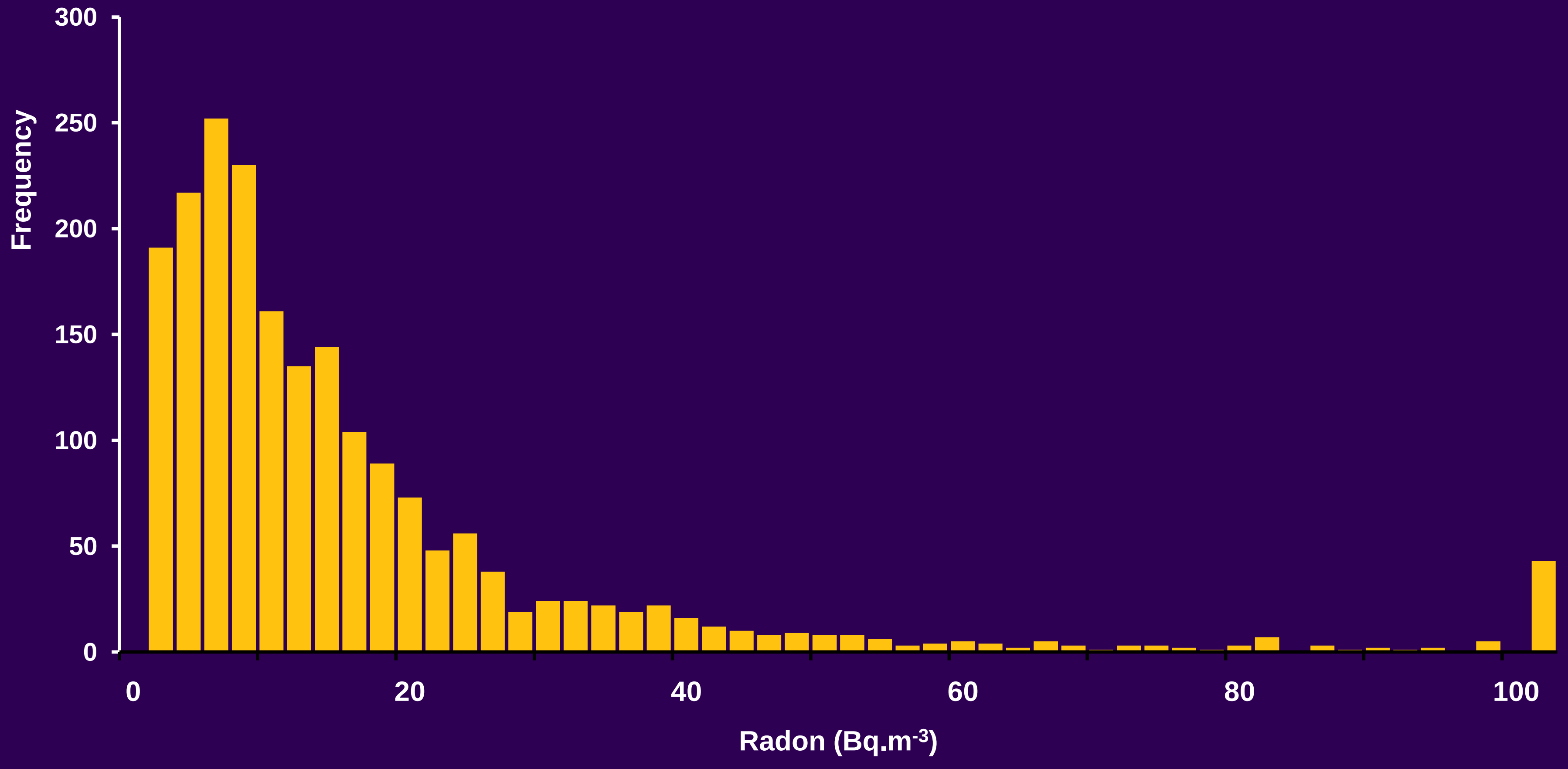
Data collected in a national survey on natural radiation exposure in UK dwellings (Wrixon et al. 1988).

Sets of etched track detectors were placed in the main living area and in the bedroom. Detectors were placed for 2 consecutive periods of 6 months.

A random sampling procedure was employed to select the representative sample of houses from the UK.

The radon concentration was measured in 2,094 homes.

Measured radon concentrations appear to follow a lognormal distribution



Lognormal distribution?

The UK distribution is not a straight line, so is not a simple lognormal distribution

Possible cause of deviation from straight line

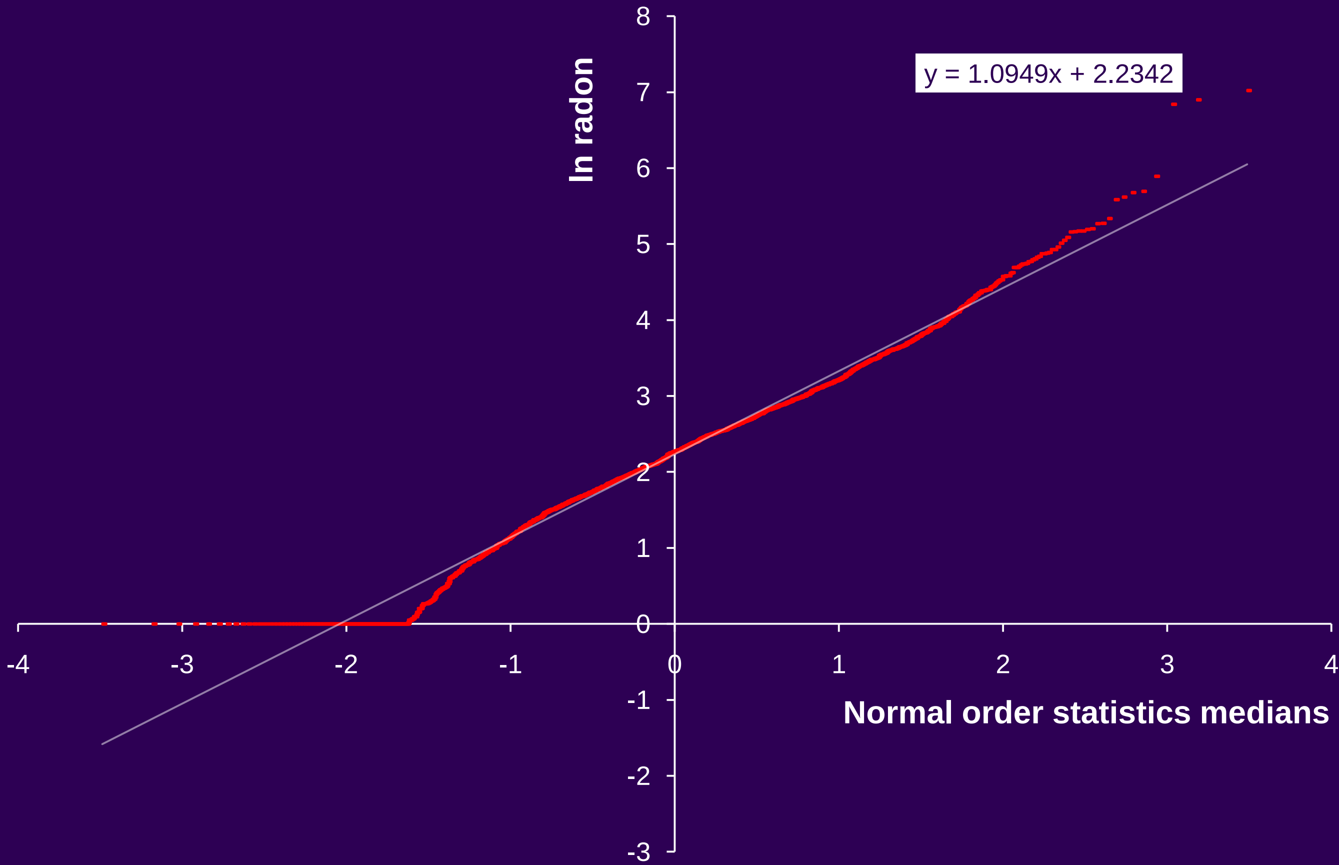
We know that the mean outdoor radon concentration in the UK is 4 Bq m^{-3}.

Outdoor radon is an addition to indoor radon which is not part of the multiplicative model for indoor radon.

Therefore we subtract 4 Bq m^{-3} from each measurement result, and replot.

The radon distribution after subtraction of outdoor radon from each result

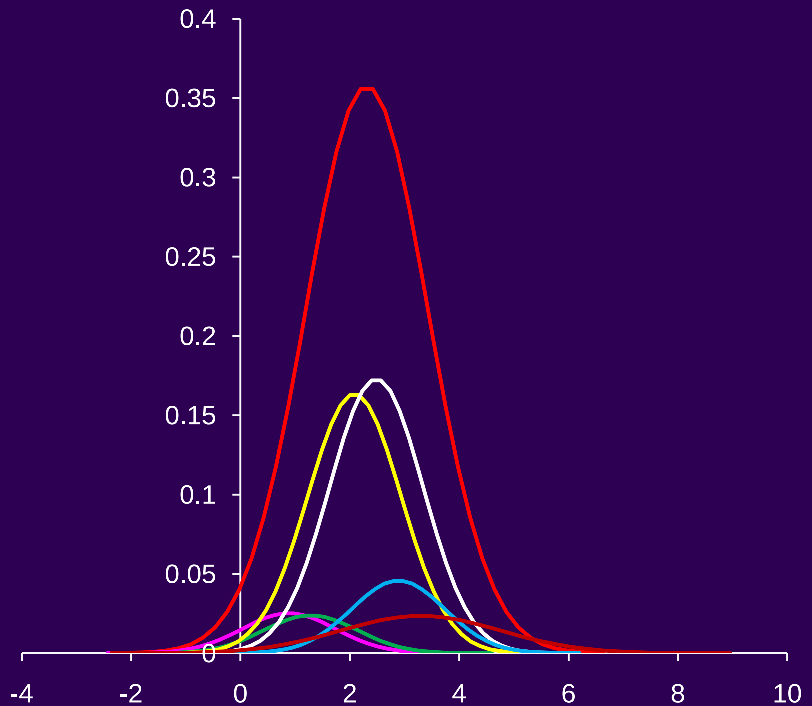
Normal probability plot of national survey data

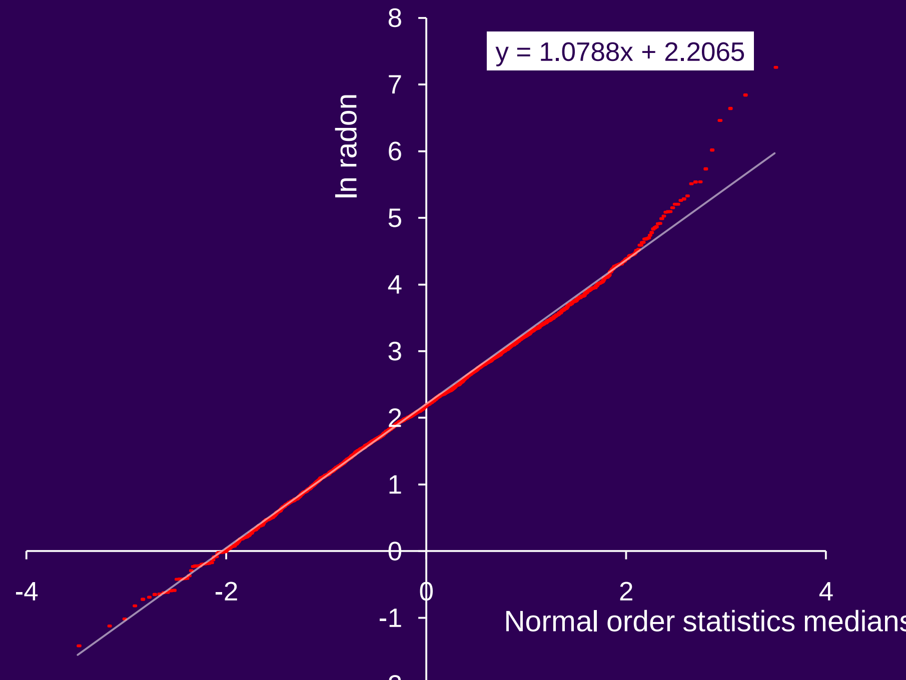


The result is a straight line over most of the range of the distribution. But there is a deviation from normality (the straight line) in the lower and upper part of the spectrum

Monte Carlo simulations of the national survey

Here we modify the 6-sample model to simulate the deviation from lognormal at high radon concentrations. The 5 samples have means of 0.8, 1.2, 2.0, 2.4, 2.8 and standard deviations of 0.85 . The 6th sample with mean of 3.2 has a standard deviation of 1.5. The samples have different weights and are not symmetrical with respect to the mean of the combined sample (red).





We therefore conclude that the deviation observed in the upper part of the national survey is probably due to the contribution from one or more local distributions with high concentrations.

Effects of random uncertainties in radon measurement results

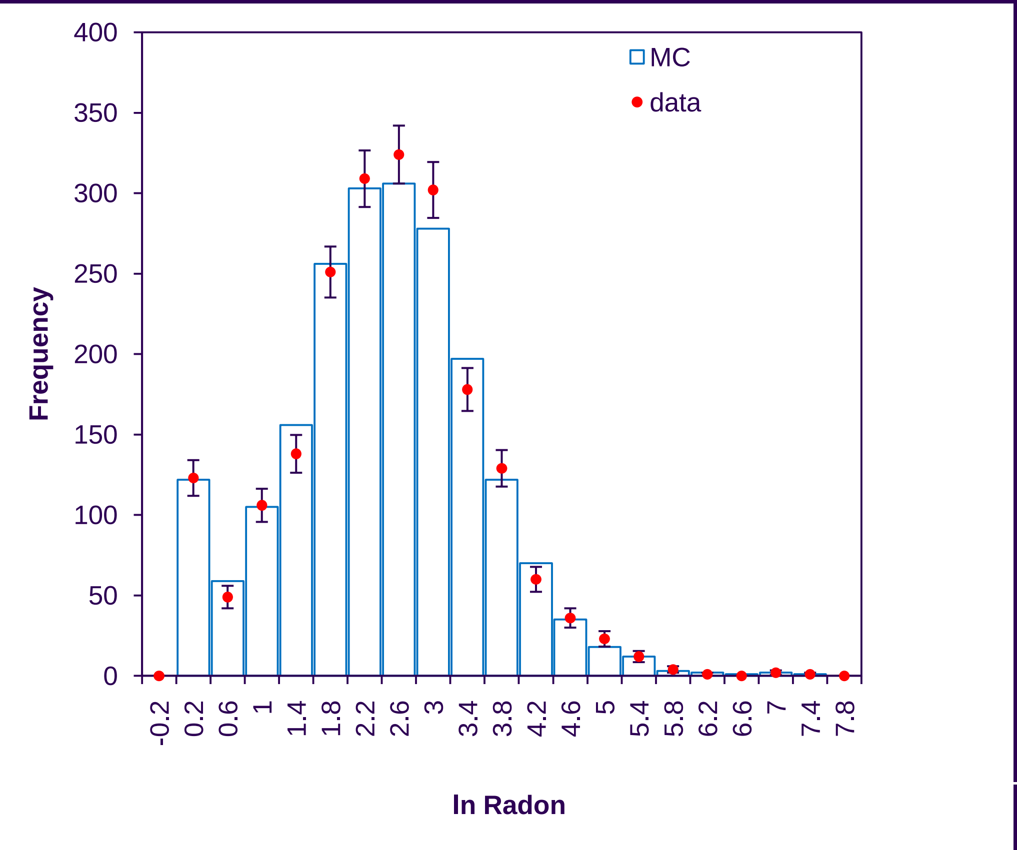
We know that there are random uncertainties in radon measurement results, and that they are proportionately larger for low radon concentrations.

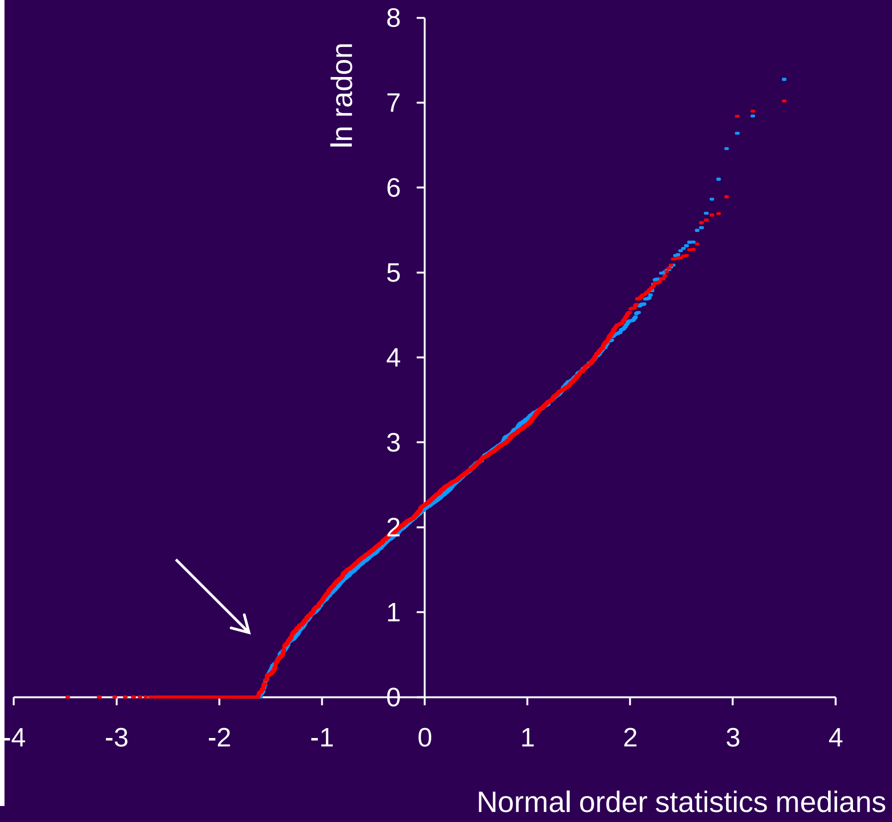
This could affect the lower end of the lognormal plot.

We will therefore add the effects of typical uncertainties in radon measurements on the simulated combined distribution on the slide 18.

Monte Carlo simulations of the national survey with measurement error

A measurement error is added to the simulations. In addition the results below zero were set to 1.





This reproduces the deviation from a straight line at low radon concentrations

Deviation from lognormal at upper end of distribution

The distribution in the national survey, if extrapolated as a straight line to higher radon concentrations, predicts that the total number of homes in the UK above a radon concentration of 5,000 Bq m^{-3} is 0.7 homes.

That was based on a systematic, unbiased survey. We have many more results from later, biased surveys.

In later surveys, we have already found 38 homes above 5,000 Bq m^{-3}, and we have certainly not yet found all of them.

Clearly the actual distribution at very high concentrations does not follow the lognormal distribution found below 200 Bq m^{-3}.

Can we use data from other surveys of UK homes to understand this deviation?

Other UK surveys

- Surveys for mapping purposes
- Surveys to identify high radon homes
- Surveys paid by homeowners
- Surveys for local authorities
- Total: > 500,000 results

BUT – All are biased, cannot be used directly to estimate changes in UK radon exposure

Use of radon house data from biased surveys

How can we use the data from the biased surveys, concentrated in high-radon areas, to supplement the data from the unbiased national survey?

Overall, based on the national survey, we estimate that there are 100,000 ± 30,000 homes in the UK above 200 Bq m^{-3}.

So far, we have found ~50,000 homes above 200 Bq m^{-3}.

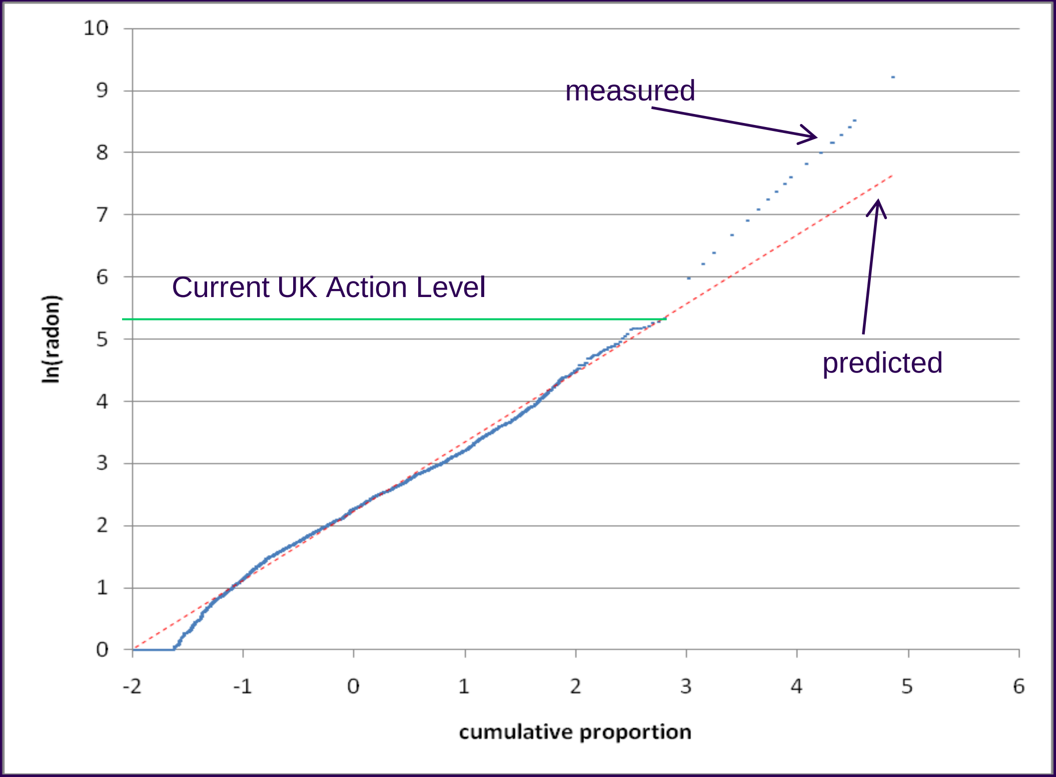
ASSUMPTION: Since we know that we have found 50% of the homes above 200 Bq m^{-3}, we will assume that we have also found 50% of those above any higher radon threshold.

Using results from biased surveys to estimate national totals

Threshold, Bq m^{-3}	Number of homes exceeding threshold		
	Estimated from national survey	Found so far	Estimated total (number found so far x2)
500	7791	9846	20000
1,000	705	2331	4500
5,000	0.7	38	75
10,000	0.02	7	15

UK distribution, with extra data

Normal probability plot, with higher estimates taken from other surveys



Here it is clear that the results at higher concentrations deviate strongly from those below 200 Bq m^{-3}, and in fact produce a second straight line, implying that the higher radon results follow a different lognormal distribution, with different mean and geometric standard deviation from the national distribution.

Implications of deviation from single lognormal distribution

- There are many more homes with very high radon concentrations in the UK than expected from the national survey
- Calculations of numbers of homes above 100 Bq m^{-3} or 200 Bq m^{-3} are not significantly changed
- Calculations for mapping are not significantly affected, since they are based on the local distribution in high radon areas
- Use of maps in surveys to find high homes is not affected

Conclusions

Health Protection Agency

- Analysis of results from the national survey in the UK has confirmed that the results follow a lognormal distribution over most of the range.
- It is necessary to allow for the contribution from outdoor radon, which is not part of the lognormal model.
- Monte Carlo simulations have shown that a lognormal distribution can be constructed from the sum of many lognormal distributions some with different weights.
- The observed distribution of national survey results can be obtained by combining several lognormal distributions, with measurement errors.
- The results above 400 Bq.m^{-3} deviate significantly from those below 200 Bq m^{-3} implying that at these concentrations the lognormal distribution has different mean and standard deviation.
- There are many more homes with very high radon concentrations in the UK than expected from the national survey, showing the importance of surveys to find and fix high homes