Did the Burglar Steal my Car Keys?

Controlling the Risk of Remains Being Missed in Archaeological Surveys

Mike Nicholson, Jon Barry and Clive Orton

1 Introduction

Consider two incidents that might trigger a fruitless search:

1. mislaid car keys, and
2. a noise in the night that might indicate a burglar.

In the first case, the search is likely to be intense and sustained. In the second, cursory and short.

There are at least two reasons why this might be. The first is our level of belief in the existence of the searched-for object. Although the car keys remain elusive, our belief in them is almost unshakeable. In the case of the burglar, (hopefully) our experience is that they are unlikely.

The second reason is related to the size of the searched-for object. A quick check should be enough to give us confidence that there is no (big) burglar. The (small) car keys however, could have been missed even during the most extensive search.

We can formalise these principles and use them to assist in designing effective surveys to detect hidden patches of objects or resources. The survey objective is to control the risk that such patches go undetected. Conversely, having carried out the survey and observed no patches, what inferences can we make about the size and frequency of patches that could have been missed?

Previous discussions about optimum design of archaeological surveys to detect patches of remains include Krakker *et al* (1983), Lightfoot (1986) and Kintigh (1988). However, the problem has arisen in many other fields, for example geology (Singer, 1975), mineralogy (Drew, 1979), pollution monitoring (Ferguson, 1992) and ecology (Nicholson and Barry, 1995).

The following section introduces the terminology and some of the basic ideas. We compare the classical with Bayesian solutions that may reduce sampling effort by incorporating prior information about the possible existence and size of remains. These results are then applied in section 3 to data from a series of excavations made in London.

2 Controlling the risk that remains of a given area are missed

Suppose that in a survey area there may be hidden (e.g. sub-surface) archaeological remains with relative area a, i.e.

$$a = \frac{\text{area of remains}}{\text{survey area}} \quad 0 \leq a \leq 1.$$

We assume that the remains are detected if one or more sampling points overlay the remains. i.e. detection does not depend on the depth or the vertical extent of the remains.

Our objective is to design the survey so that if there are remains with relative area equal to or greater than some value a^*, the risk that they will be missed is sufficiently small.

We begin by discussing the classical solution to this problem. We then go on to describe two Bayesian models (Hartigan, 1983; Lee, 1989), the first incorporating prior information about the existence of remains, and the second incorporating information about both existence and the size of remains. Finally we consider the changes to the measured risk brought by this additional information.

We derive simple, explicit results for simple random surveys of N independent sampling points – i.e. the outcome at one sampling point isn't affected by the outcome at other sampling points. The implications of other sampling schemes are presented in the discussion.

2.1 Classical solution

For a random survey of N independent sampling points, the relative area of remains is equivalent to the probability of success in a binomial distribution of the number of successes in N trials. When no remains are detected, the upper (1- p) confidence limit for a (Johnson and Leone, 1977, p267) is defined as a_p, the largest a satisfying

$$1\text{-}p \leq (1 - a)^N.$$

Turning this around, we can therefore define $(1\text{-}a^*)^N$ as a upper limit of the (classical) risk that remains with a relative area greater than or equal to a^* have been missed i.e.

$$R_{classical}(a \geq a^*) \leq (1 - a^*)^N.$$

For example, suppose that the total survey area is 100x100 m^2. If there was no evidence of any remains at $N = 50$ random sampling points, the risk that remains with an area of 100 m^2 (relative area = 1%) or greater could have been missed is

$$R_{classical}(a \geq 0.01) \leq (1 - 0.01)^{50} = 0.61.$$

We see that even with 50 sampling points, this risk is very high – worse than the toss of a coin. This reflects the difficulty of making inferences about small probabilities with limited resources. Table 1 demonstrates this further, using

$$N = \frac{\log[R_{classical}(a \geq a^*)]}{\log(1 - a^*)}$$

to find the values of N required for different combinations of $R(a \geq a^*)$ and a^*.

Note also that for small a^*,

$$(1 - a^*)^N \approx e^{-a^* N}$$

showing that approximately, a given risk effectively depends on the product of a^* and N. This implies that e.g. halving a^* (by decreasing the area of the remains or increasing the survey area) requires an approximate doubling of N.

2.2 Bayesian Model 1: Incorporating information about the likely existence of remains of a given size

Table 1 demonstrated that for small remains in large areas, relying solely on surveys to control risk demands large sampling resources. It may be possible to reduce this demand however, by adopting a Bayesian approach to incorporate additional information. In this first model, we will incorporate prior information about the likely existence of remains.

Suppose that there may be remains with relative area a^* or greater in a survey area, but that no remains are detected in a survey. Before the survey, we can quantify our prior belief as π_{prior}, the prior probability that these remains exist. We can then update π_{prior} having carried out the survey.

Let $\pi_{posterior}$ be the posterior belief that there are remains with area greater than or equal to a^* in the survey area given that no remains were seen at N sampling points. Deriving $\pi_{posterior}$ is simply an exercise in conditional probability:

i.e.

$$\Pr(\text{remains exist} \mid N \text{ missed}) = \frac{\Pr(\text{remains exist and } N \text{ Missed})}{\Pr(N \text{ missed})}$$

$$= \frac{\Pr(\text{remains exist}) \times \Pr(N \text{ missed} \mid \text{remains exist})}{\Pr(\text{remains don't exist}) + \Pr(\text{remains exist}) \times \Pr(N \text{ missed} \mid \text{remains exist})}$$

i.e.

$$\pi_{posterior} = \frac{\pi_{prior} \mathrm{M}_{a*}}{1 - \pi_{prior} + \pi_{prior} \mathrm{M}_{a*}}$$

where M_{a*} is the probability that remains of relative area a^* are missed at all N sampling points. For a simple random sample

$$\mathrm{M}_{a*} = (1 - a^*)^N .$$

Clearly π_{prior} and $\pi_{posterior}$ correspond to the prior and posterior risks that remains of relative area a^* or greater could have been missed in the survey, giving

$$R_{prior}(a \geq a^*) = \pi_{prior}$$

and

$$R_{posterior}(a \geq a^*) = \pi_{posteriior} = \frac{\pi_{prior}(1 - a^*)^N}{1 - \pi_{prior} + \pi_{prior}(1 - a^*)^N} .$$

When designing a survey, this last equation can be re-written to give

$$N = \frac{\ln\left(\dfrac{\pi_{posterior}}{\pi_{prior}} \dfrac{(1 - \pi_{prior})}{(1 - \pi_{posterior})}\right)}{\ln(1 - a^*)} ,$$

the required number of sampling points in a simple random sample required to reduce the risk of failing to discover remains with relative area a^* or greater from π_{prior} to $\pi_{posterior}$. For example, Figure 1 shows $R_{posterior}(a \geq 1\%)$ plotted against $R_{prior}(a \geq 1\%)$ for N = 25, 50, 100, 250 and 500.

Differences between the sample size requirements suggested by this approach and the classical solution are discussed in Section 2.4. But first we consider an extended version of this Bayesian model where there is prior information about both the existence of remains and their size.

2.3 Bayesian Model 2: Incorporating prior information about both the likely existence and size of remains

In this second case, we model existence and size of remains separately. For existence, we re-define π_{prior} and $\pi_{\text{posterior}}$ to be the prior and posterior probabilities that a survey area contains any remains – i.e. remains of any size.

In the second component, the relative area of the remains, a, is now a random variable, whose variation is described by some probability distribution. Prior information about area is expressed in the parameters of this distribution. Let us deal with second component first.

A convenient probability distribution to consider is a specific parameterisation of the Beta distribution

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

where $B(\alpha,\beta)$ is the mathematical beta function (Evans *et al*, 1993). Following Nicholson and Barry (1995), we set $\alpha = 1$. The prior distribution for a is then

$$f(a) = (1-a)^{\beta_{prior}-1}\beta_{prior}$$

and the mean of a is given by

$$E[a] = \frac{1}{1+\beta_{prior}}.$$

Hence prior information about the area of remains, given they exist, can be expressed simply as the average relative area. Alternatively, as we shall see, β_{prior} acts as a sort of virtual sample size – the number of observations in a hypothetical previous survey in which nothing was found (c.f. Nicholson and Barry, 1995).

Prior Distribution

To obtain the prior distribution function for the relative area of remains, we combine the probability of existence and the distribution of relative area where there are remains as follows:

write

Pr($a < a^*$) = Pr(no remains, or, remains exist and $a < a^*$)

or

Pr($a < a^*$) = Pr(no remains) + Pr(remains exist) × Pr($a < a^*$ | remains exist).

i.e. the cumulative prior distribution function of the relative area of remains, that may or may not exist is given by

$$H(a^*) = 1 - \pi_{prior} + \pi_{prior} \times F(a^* \mid \beta_{prior})$$

where $F(a^*|\beta_{prior})$ is the cumulative distribution function for a Beta probability density with parameters $\alpha = 1$ and β_{prior}.

Posterior Distribution

Suppose again that we carry out a random survey of N sampling points, at which no remains are observed. To find the corresponding posterior distribution, $G(a^*)$, we have

$G(a^*) = \Pr(a<a^*|\text{No remains seen})$

which is derived in Appendix 1. The solution is

$$G(a^*) = 1 - \pi_{posterior} + \pi_{posterior} \times F(a^* \mid \beta_{posterior})$$

where

$$\pi_{posterior} = \frac{\pi_{prior} \times {\beta_{prior}} \big/ {N + \beta_{prior}}}{1 - \pi_{prior} + \pi_{prior} \times {\beta_{prior}} \big/ {N + \beta_{prior}}}$$

and

$$\beta_{posterior} = N + \beta_{prior} \,.$$

We see that $H(.)$ and $G(.)$ are of the same form.

Prior and Posterior Risk

The prior and posterior risks are simply given by

$$R_{prior}(a \geq a^*) = 1 - H(a^*) = \pi_{prior}(1 - a^*)^{\beta_{prior}}$$

and

$$R_{posterior}(a \geq a^*) = 1 - G(a^*) = \pi_{posterior}(1 - a^*)^{\beta_{posterior}} \,.$$

Samples sizes required for a given posterior risk must obtained numerically. For example, Figure 2 shows $R_{prior}(a\geq 1\%)$ for a prior mean relative area of 2% (equivalent to β_{prior} =49) plotted against π_{prior}, together with the corresponding $R_{posterior}(a\geq 1\%)$ for N = 25, 50, 100, 250 and 500.

Predictably, Figure 2 shows the posterior risk generally increasing with increasing π_{prior}. However, the behaviour of the posterior risk for changing mean area of remains is less straightforward. Figure 3 shows $R_{prior}(a\geq 1\%)$ for π_{prior} = 0.2 plotted against the prior relative mean area of remains together with the corresponding $R_{posterior}(a\geq 1\%)$ for N = 25, 50, 100, 250 and 500. We see that although the posterior risk decreases with increasing N, it first increases and then decreases with increasing prior mean area of remains.

With more thought, this is sensible – since if the area of remains is high where they exist, the less likely they are to exist when none have been seen in the survey. Figures 4 and 5 support this simple argument by showing how the components of the posterior risk change with the prior mean area of remains. Figure 4 shows the posterior mean area of remains increasing with the prior mean, whereas Figure 5 shows $\pi_{posterior}$ decreasing with increasing prior mean. Together, these opposing responses produce the complex response seen in Figure 3.

2.4 Comparisons of Classical and Bayesian Risks

Classical Versus Bayesian model 1

Comparing $R_{posterior}(a \geq a^*)$ for model 1with $R_{classical}(a \geq a^*)$, we find

$R_{posterior}(a \geq a^*) > R_{classical}(a \geq a^*)$

when

$$\pi_{prior} > \frac{1}{2-(1-a^*)^N} .$$

Thus with e.g. $a^* = 1\%$ and $N = 100$, when the prior evidence for the existence of remains is high, i.e.

$$\pi_{prior} > \frac{1}{2-0.99^{100}} = 0.612 ,$$

the posterior risk will be higher than the classical risk. i.e. N = 100 is insufficient to offset such a high initial probability that there are remains.

However, having prior information about the possible existence of remains will give a smaller posterior risk when

$$\pi_{prior} < \frac{1}{2-(1-a^*)^N} .$$

Since this is satisfied for any a^* and N when

$\pi_{prior} < 0.5$,

having prior information about the existence of remains will lead to a smaller risk than the classical model provided non-existence is more likely than existence.

Classical Versus Bayesian model 2

Comparing $R_{posterior}(a \geq a^*)$ from model 2 with $R_{classical}(a \geq a^*)$, we have

$$R_{posterior}(a \geq a^*) = \pi_{posterior}(1-a^*)^{\beta_{posterior}}$$

or

$$R_{posterior}(a \geq a^*) = \pi_{posterior}(1-a^*)^{N+\beta_{prior}} = \pi_{posterior}(1-a^*)^{\beta_{prior}} \times R_{classical}(a \geq a^*) .$$

Hence

$$R_{posterior}(a \geq a^*) < R_{classical}(a \geq a^*)$$

for any $a^*>0$ and $\beta_{prior}>0$.

3 Application to archaeological sites in London

To demonstrate how these results might be applied, we use data on the frequency and area of archaeological remains reported for Greater London. Information about the area of remains comes from published plans of archaeological sites, of which the most abundant were from Roman sites. We therefore initially assess the risk of failing to detect Roman remains, and then generalise these results to consider any type of remains.

Table 2 summarises the frequency of positive and negative evaluations of sites assessed in 1992 and 1993 in the thirty-three London boroughs and other local authorities which make up Greater London (McCracken and Phillpotts, 1995). From the total of 120 positive evaluations out of 414, we have an estimate of π_{prior}, the prior probability that there are remains at a site, of 0.29.

Table 3, also taken from McCracken and Phillpotts (1995), summarises the frequency of remains from different periods. Combined with the overall value of π_{prior}, these frequencies then give the prior probabilities of remains from a specific period; e.g. the value of π_{prior} for Roman remains is 0.29*30/206 = 0.042.

Information about the area of remains comes from the published plans of five sites containing Roman constructions (Table 4). From this admittedly small data set, an estimate of the prior mean of the relative area of remains is 12.5%, corresponding to $\beta_{prior} = 7$.

Putting this information together into Bayesian model 2 (Section 2.3), the prior risk that e.g. remains with a relative are of 1% or greater could be present at a site is

$$R_{prior}(a \geq 1\%) = \pi_{prior}(1-0.01)^{\beta_{prior}} = 0.039\,.$$

Figure 6 shows the corresponding posterior risk that Roman remains with a relative area greater that 1% could have been missed in a random survey as a function of N, the number of sampling points. We see $R_{posterior}$(a>1%) =1/10 would require $N = 24$, whereas $R_{posterior}$(a>1%) =1/100 would require $N = 115$.

Without more information, it is difficult to extend these results to remains from any period. Although information about the areas of non-Roman remains was less easy to obtain, we note that in some cases, non-Roman remains overlaid or were overlaid by Roman remains (c.f. Adkins and Adkins, 1983; Williams, 1984; Potter, 1994). Hence we might argue that the average relative area derived from Table 4 is appropriate to remains from any period. Alternatively we might simply employ Bayesian Model 1 (Section 2.2) and assume we have no information about the distribution of the size of remains from any period. Figure 7 shows the prior and posterior risks for both of these approaches.

Table 5 summarises and compares the sample sizes for the classical and Bayesian models corresponding to risks of 1/10, 1/100, 1/1000 and 1/10000 for both Roman and any remains when no remains have been seen in the survey.

4 Discussion and conclusions

The results presented here show the large potential for decreasing sampling requirements when prior information about the existence and size of remains is available.

We have dealt with the simple case where detection is certain if the remains overlay one or more sampling points. Reality may require more complex assumptions. For example, Krakker *et al.* (1983) and Kintigh (1988) considered uncertain detection in terms of the density and visibility of artifacts within a site. However, the simple results presented here might still be useful. Intuitively, suppose artefact density, observer efficiency or other factors result in a probability $p_{detection}$ that remains are detected when hit by a sampling point. Then the effect will simply be to re-scale the area of remains from a to an effective area $p_{detection} \times a$.

Kintigh (1988) and Lightfoot (1986) also considered the effect of different size of sampling units. Again intuitively, the effect will be to re-scale the area of the remains – effectively extending the remains with a surrounding band whose width depends on the size and shape of the sampling unit.

We have also only considered a random sampling design. This leads to simple results that are applicable to remains of any shape, present in any number of fragments. Many authors have considered sampling schemes that are potentially more efficient. For example, Singer (1975), Drew (1979), Krakker *et al.* (1983), Kintigh (1988) and Ferguson (1992) have discussed and compared the effectiveness of various systematic sampling schemes, in particular square- and triangular lattice designs. However, the effectiveness of these designs depends on the shape and number of targets. Drew (1979) developed theoretical results for a fixed number of elliptical targets. Ferguson (1992) also considered a fixed number of targets, but derived results by simulation for several simple geometrical shapes. Nicholson and Barry (1999) derived theoretical results for a random number of circular targets, with frequency generated by a Poisson distribution. They employed the Bayesian approach described here to provide posterior estimates of the number and average radius of circular targets.

However, for remains with a complex shape, particularly when the total area of remains is fragmented, the performance of lattice designs and simple random sampling may be very similar. For example, Figure 8 shows the probability that remains present as a single circular target go unseen with N = 50 sampling points for both simple random sampling and (the most efficient) triangular lattice design. These were computed using the formulae of Barry and Nicholson (1993), who give exact formulae for several designs for a single circular target. For circular remains with an area of about 2%, detection is almost certain for the triangular design, but there is a 36% chance they will be missed with the random sample. Also shown is the effect of these remains being present as a number of randomly dispersed, equally sized circular fragments. We see that the benefits of the triangular design are quickly lost when fragmentation occurs. Hence if complex-shaped or fragmented remains are possible, the results presented here may be adequate for both regular and random designs. At worst, they would provide an upper limit for the posterior risk where no remains have been seen. Note that the Roman remains summarised in Table 4 tended to be very fragmented.

We have presented results that allow us to quantify the risk that remains could be, or have been, missed. An interesting question that has not been discussed is what inferences can be made if one or more sampling points do encounter some remains. With random sampling, the number of hits divided by N provides a simple estimate of the total area of remains. However, total area may not be very informative, if hits from two contiguous sampling points indicate potentially larger and more-significant remains than two unconnected sampling points.

One approach might be for evidence of remains from a few sampling points to trigger different decisions, such as to invoke further, more extensive sampling. Thompson and Seber, 1996 describe adaptive sampling strategies, where resources in a second stage of sampling are focussed on areas of interest suggested by the first stage.

Note that the Bayesian approach suggested here naturally lends itself to this multi-stage sampling approach: the posterior values of the model parameters simply become the prior values in subsequent waves of sampling. From the form of $R_{posterior}$ for model 2, we see that the role of β_{prior} is simply to kick-start this process, acting as a sort of virtual survey, providing an initial estimate of risk. A practical strategy for dealing with a large number of sites might be to screen all sites with a low sampling intensity to reduce this initial risk and flush out sites with large remains. Sites with no evidence of remains can then be ranked according to their posterior risks. Both sites with evidence of remains and high-risk sites might then be subject to further sampling or evaluation.

Another refinement might be to compute π_{prior} on a regional basis e.g. separately for each of the thirty-three London Boroughs. Alternatively, it may be possible to model π_{prior} as a function of soil type, land use, and so on. This would lead to prior risks based on local knowledge, and greater control of the resulting risks that archaeological remains are lost.

References

Adkins, L. and Adkins, R. (1983) Excavations at Beddington 1982. *London Archaeologist* **4** 326−329.

Barry, J. and Nicholson, M.D. (1993) Measuring the probability of patch detection for four spatial sampling designs. *Journal of Applied Statistics* **20** 353−362.

Drew, L.J. (1979) Pattern drilling explorations: Optimum pattern types and hole spacings when searching for elliptical shaped targets. *Mathematical Geology* **11** 223−254.

Evans, M., Hastings, N. and Peacock, B. (1993) *Statistical Distributions* (Second edition) John Wiley and Sons, New York.

Ferguson, C.C. (1992) The statistical basis for spatial sampling of contaminated land. *Ground Engineering* **25** 34−38.

Hartigan, J.A. (1983) *Bayes Theory* Springer-Verlag, New York.

Johnson, N.L. and Leone, F.C. (1977) *Statistics and Expermental Design in Engineering and the Physical Sciences* Vol. 1. 2nd ed. Wiley, New York.

Kintigh, K.W. (1988) The effectiveness of subsurface testing: a simulation approach. *American Antiquity* **53** 686−707.

Krakker, J.J., Shott, M.J. and Welch, P.D. (1983) Design and evaluation of shovel-test sampling in regional archaeological survey. *Journal of Field Archaeology* **10** 469−480.

Lee, P.M. (1989) *Bayesian Statistics: an Introduction.* Edward Arnold, Sevenoaks, Kent.

Lightfoot, K.G. (1986) Regional surveys in the eastern United State: the strengths and weaknesses of implementing subsurface testing programs. *American Antiquity* **51** 484−504.

McCracken, S. and Phillpotts, C. (1995) Archaeology and planning in London. Assessing the effectiveness of PPG 16. *Standing Conference on London Archaeology* **10** 469−480.

Miller, L. (1982) Miles Lane: the early Roman waterfront. *London Archaeologist* **4** 143−147.

Nicholson, M.D. and Barry, J. (1995) Inference from spatial surveys about the presence of an unobserved species. *Oikos* **72** 74−78.

Nicholson, M.D. and Barry, J. (1999) Inference from spatial surveys about the distribution of patch size of an unobserved species. *The Statistician* **48** 327−337.

Potter, G. (1994) 15-17 Brighton Road, Croydon: the investigation of a prehistoric and Roman site. *London Archaeologist* **7** 232−237.

Rowsome, P. (1996) The Billingsgate Roman house and bath – conservation and assessment. *London Archaeologist* **7** 415−423.

Singer, D.A. (1975) Relative efficiencies of square and triangular grids in the search for elliptically shaped resource targets. *Journal Research US Geological Survey* **3** 163−167.

Thompson, S.K. and Seber, G.A.F. (1996) *Adaptive Sampling*. John Wiley and Sons, New York.

Williams, T. (1984) Excavations at 25-26 Lime Street. *London Archaeologist* **4** 426–430.

Appendix 1

$G(a^*) = \Pr(a<a^*|N \text{ missed})$

$$= \frac{\Pr(a < a^* \; and \; N \text{ missed})}{\Pr(N \text{ missed})}.$$

Now,

$$\Pr(N \text{ missed}) = 1 - \pi_{prior} + \pi_{prior} \int_0^1 (1-a)^N (1-a)^{\beta_{prior}-1} \beta_{prior} da$$

$$= 1 - \pi_{prior} + \pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}}.$$

And

$$\Pr(a < a^* \; and \; N \text{ missed}) = 1 - \pi_{\text{prior}} + \pi_{prior} \int_0^{a^*} (1-a)^N (1-a)^{\beta_{prior}-1} \beta_{prior} da$$

$$= 1 - \pi_{prior} + \pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}} \left\{1 - (1-a^*)^{N+\beta_{prior}}\right\}$$

Hence

$$G(a^*) = \frac{1 - \pi_{prior} + \pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}} \left\{1 - (1-a^*)^{N+\beta_{prior}}\right\}}{1 - \pi_{prior} + \pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}}}$$

$$= \frac{1 - \pi_{prior}}{1 - \pi_{prior} + \pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}}} + \frac{\pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}}}{1 - \pi_{prior} + \pi_{prior} \frac{\beta_{prior}}{N + \beta_{prior}}} \left\{1 - (1-a^*)^{N+\beta_{prior}}\right\}$$

or

$$G(a^*) = 1 - \pi_{posterior} + \pi_{posterior} \times F(a^* \mid \beta_{posterior})$$

where

$$\pi_{posterior} = \frac{\pi_{prior} \times {\beta_{prior}}/{N + \beta_{prior}}}{1 - \pi_{prior} + \pi_{prior} \times {\beta_{prior}}/{N + \beta_{prior}}}$$

and

$$\beta_{posterior} = N + \beta_{prior}.$$

Table 1 Sample size N for different combinations of $R_{classical}(a \geq a^*)$ and a^*.

$R_{classical}(a \geq a^*)$		a^*	
1 in:	10%	1%	0.1%
10	22	230	2302
100	44	459	4603
1000	66	688	6905
10000	88	917	9206

Table 2 Frequencies of positive and negative evaluations for Greater London. Source: McCracken and Phillpotts (1995).

Year	No. of Sites	No. Positive	No. Negative
1992	188	54	134
1993	226	66	160
Total	414	120	294

Table 3 Frequencies of positive evaluations by period of remains. Source: McCracken and Phillpotts (1995).

Year	Pre Historic	Roman	Saxon	Medieval	Post Medieval	Total
1992	29	15	2	14	40	100
1993	43	15	5	13	30	106
Total	72	30	7	27	70	206
π_{prior}	0.10	0.042	0.009	0.038	0.098	0.29

Table 4 Areas of Roman remains.

Source	Description	Area of remains %
Miller (1982)	Roman waterfront	8.6
Adkins and Adkins (1983)	Roman buildings	17.1
Williams (1984)	Roman buildings	6.6
Potter (1994)	Roman features	15.2
Rowsome (1996)	Roman house and bath	15.0
	Average area	12.5

Table 5 Summary of sample sizes *N* required to control the risk that remains with an area of 1% would be missed at N random sampling points.

	Roman remains	Any remains		
Posterior Risk	Model 2	Model 2	Model 1	Classical
	$\pi_{prior} = 0.056$	$\pi_{prior} = 0.29$	$\pi_{prior} = 0.29$	
1 in:	$\beta_{prior} = 7$	$\beta_{prior} = 7$		
10	-	14	130	229
100	17	94	369	450
1000	100	237	599	688
10000	243	413	828	917

Figure 1

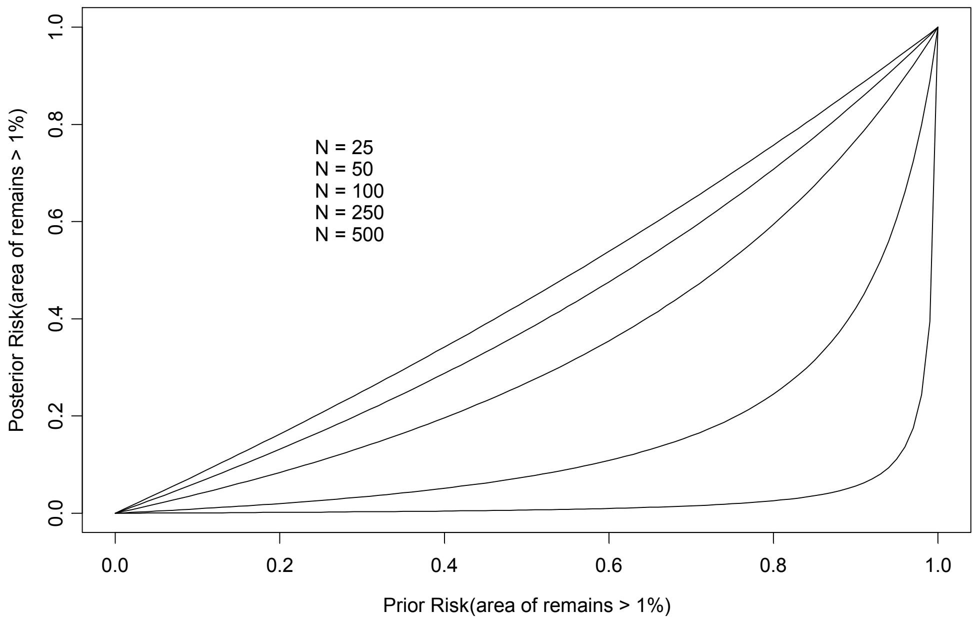

N = 25
N = 50
N = 100
N = 250
N = 500
Posterior Risk(area of remains > 1%)
Prior Risk(area of remains > 1%)
0.0
0.2
0.4
0.6
0.8
1.0

Figure 2

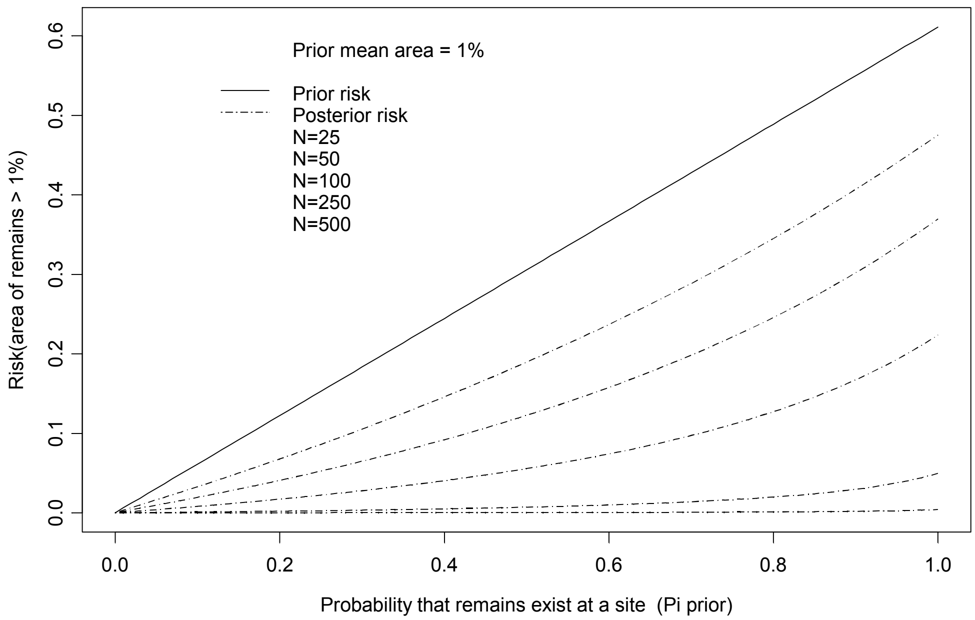


Figure 3

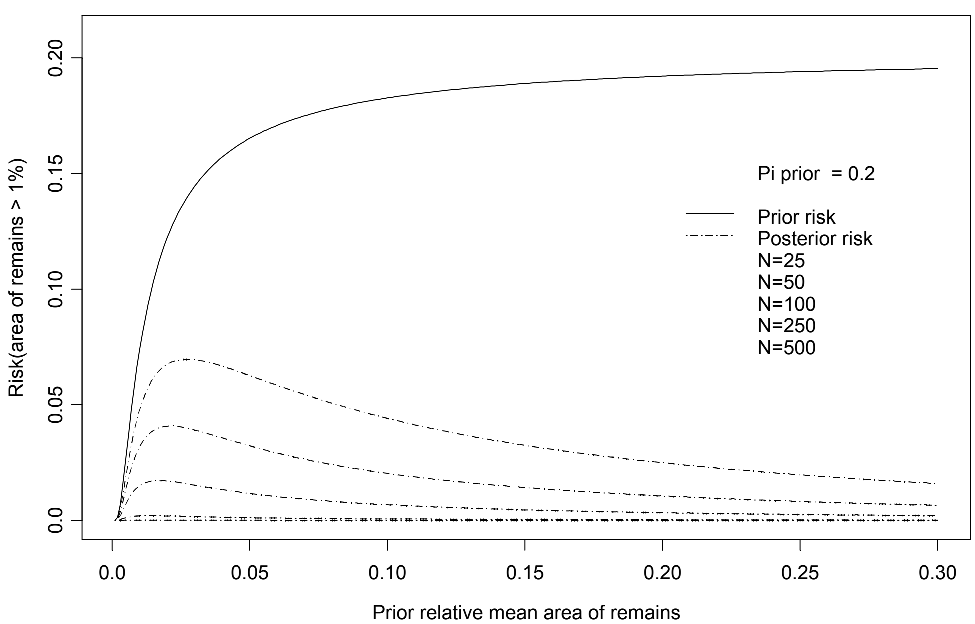


Figure 4

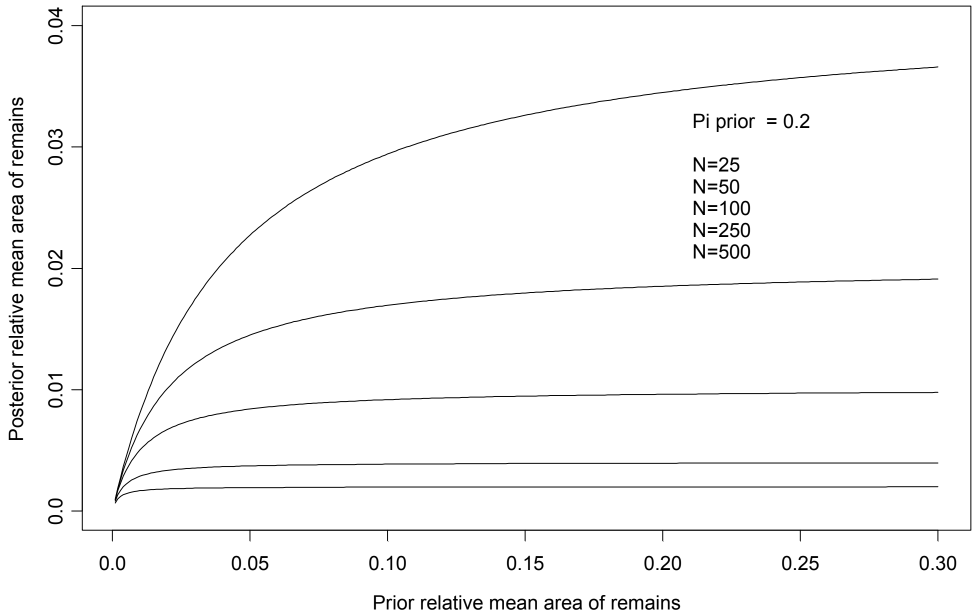


Figure 5

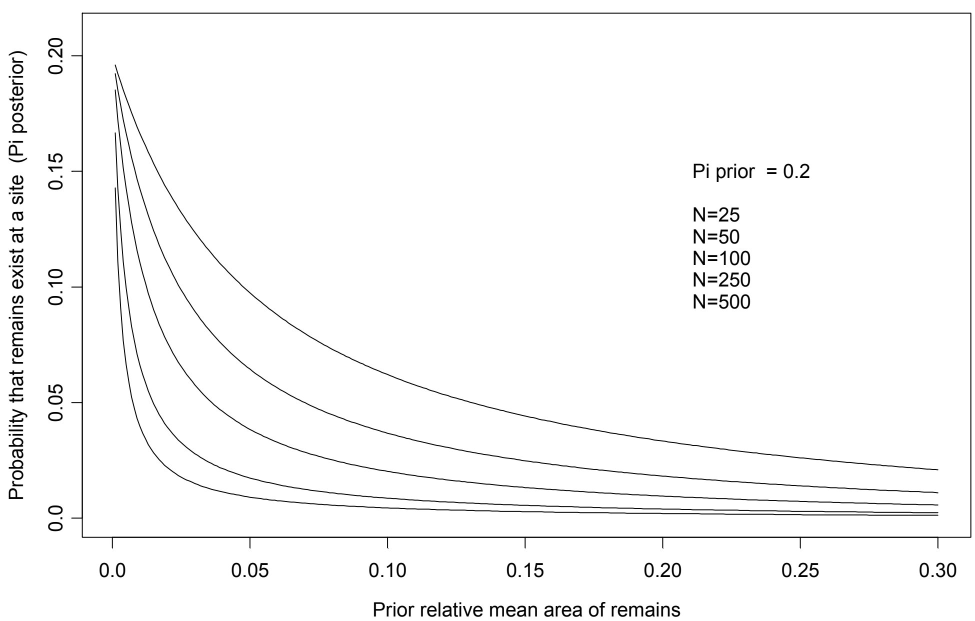


Figure 6

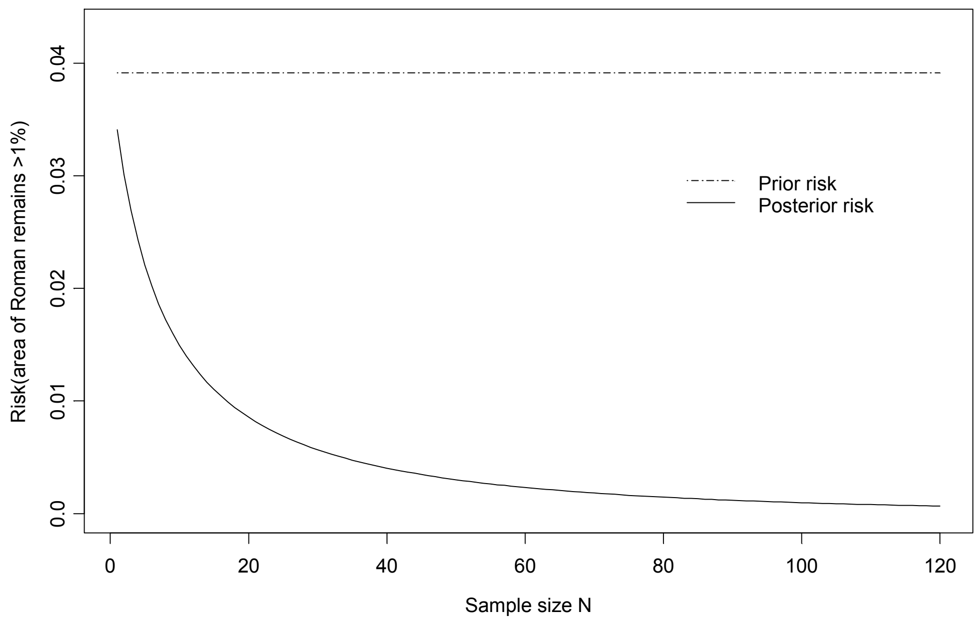


Figure 7

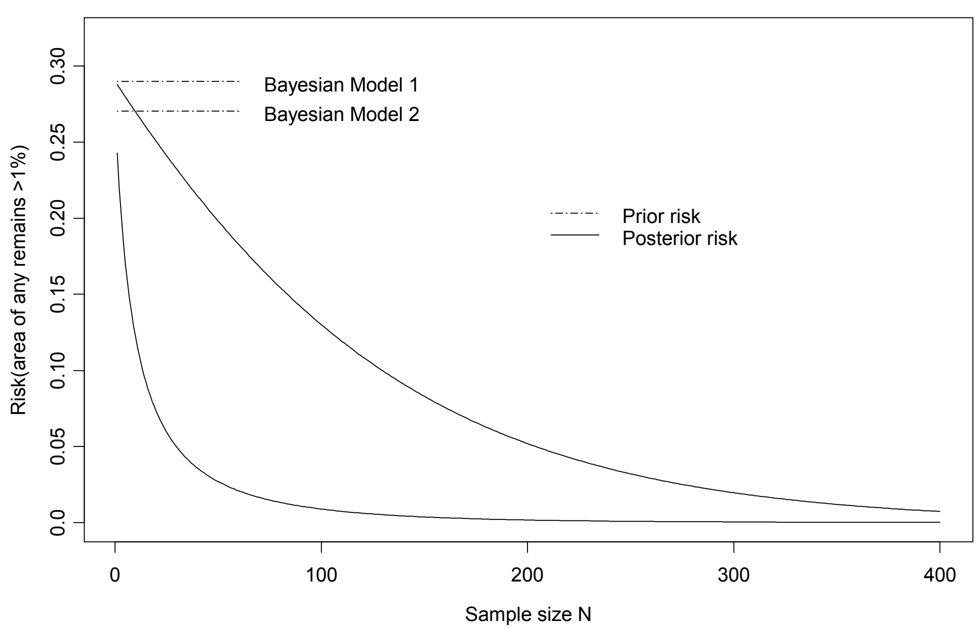
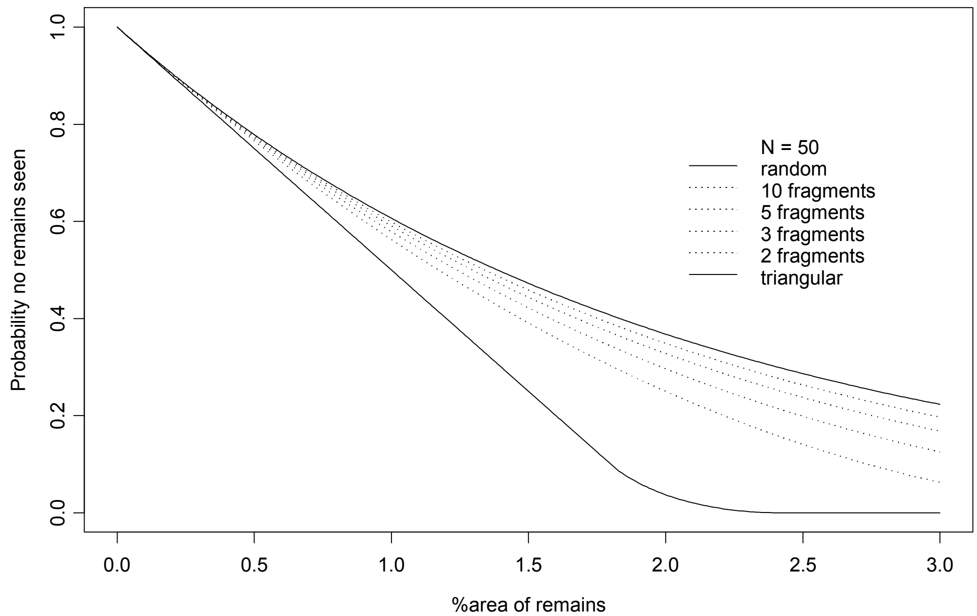


Figure 8


N = 50
random
10 fragments
5 fragments
3 fragments
2 fragments
triangular
Probability no remains seen
%area of remains
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0