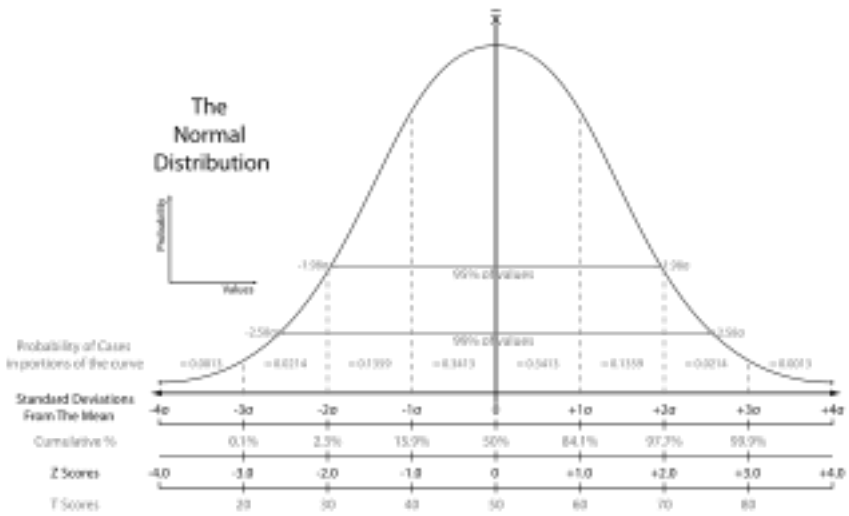
Is it safe to go out yet?

Statistical inference in a zombie outbreak model



Modelling

- Mathematical modelling describes how physical systems change over time
- These include laws of motion
  - planetary orbits, ballistic missiles, elastic springs, chemical reactions, radioactive decay
- Modelling involves
  - identifying the essential features of a physical system
  - converting these into a descriptive mathematical framework
  - making predictions.

Ordinary Differential Equations

- ODEs can be effective at characterising the main features of a system
- Even when there are a range of complicated objects interacting in ways that are impossible to pin down
- ODEs are the “gateway drug” between the real world and mathematics.

The power of models

- The aim is to
  - *explain* the key mechanisms at work and how they interact
  - *predict* the future state of the system
  - theoretically and computationally *investigate* "what if" scenarios
- The latter may be out of reach of experimental scientists due to
  - ethical considerations
  - budgetary limitations
  - technological feasibility.

A specific example

- Soldiers patrolling the are have reported daily observed zombie numbers for the past 5 days as 123, 127, 104, 92 and 74
- Is it safe to go out yet?
- If you meet a zombie, what is your chance of fighting it off?
- How many soldiers should be mobilise?
  - how many of these will survive?
- What scale of quarantine would be worthwhile?
- How effective does a cure need to be?

Parameters

- Mathematical modelling can make predictions in the absence of data
- However, these predictions are often quite general
- To make the model really useful, we must *calibrate* any unknown parameters
- We do this using observed data
- We may also wish to choose the best model from among a choice of potential models
- This is known as *model selection*.

Incompatible rumours

- Suppose we hear incompatible rumours that
  - zombies only attack alone
  - zombies always attack in pairs
- These two scenarios correspond to different ODE models
- To investigate which rumour is more likely to be true, we could ask which of two models best explains the observed data
- This requires us to simultaneously calibrate and compare two or more ODE models.

A simple model

- Let S(t) and Z(t) denote the concentration levels of humans and zombies at time t
  - they will take real values, so they aren't necessarily whole numbers
- The only event that can cause a change in these levels is a successful zombie attack
- We introduce a rate constant β that characterises the ability of zombies to find and infect humans
- The larger β, the more virulent the zombies
- This is called mass-action transmission.

Simple model equations

- The simple model is then

$$S'(t) = -\beta S(t) Z(t)$$
$$Z'(t) = \beta S(t) Z(t)$$

- Adding the two equations together gives (S(t)+Z(t))′=0
- Thus, the total population remains constant
  - this is not true in general
- Thus, we have S′(t)=-βS(t)(K-S(t)), so

$$S(t) = \frac{S(0)K}{S(0) + (K - S(0))\, e^{\beta K t}}.$$

S: humans Z: zombies
β: transmissibility

Known and unknown parameters

- Suppose that the size of the population K is known
- And the initial number of humans S(0) is also known
- Then β is the only unknown parameter.
- This can be interpreted as the rate at which humans are converted into zombies
- Units are “per zombie per day”.

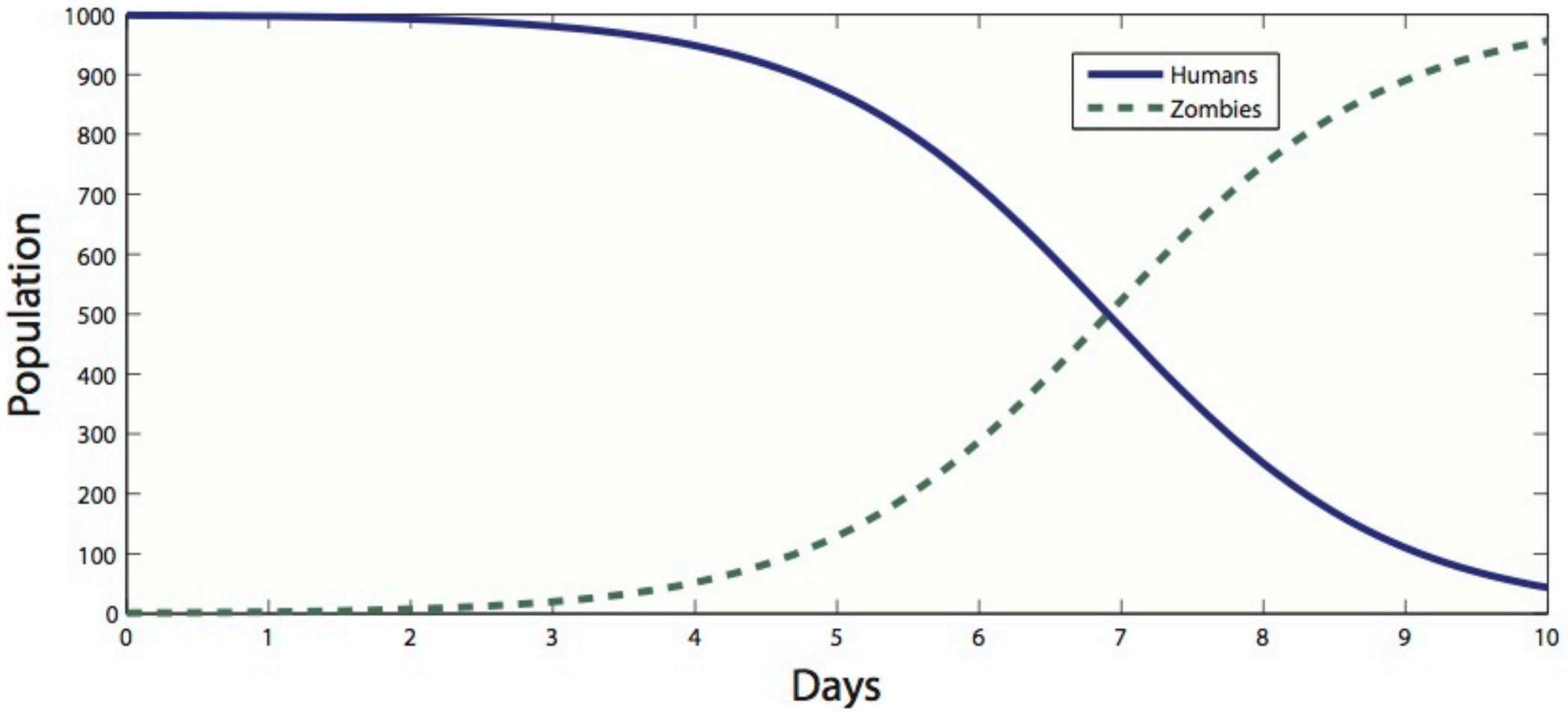
S: humans K: population
β: transmissibility

Determining β

- If there are 1000 humans, then a rate of β=0.001 initially corresponds to
- S(0)×β=1000×0.001=1 human being converted by each zombie per day
- The number of humans being converted depends on both the number of humans and the number of zombies at any particular time
- By solving the ODEs, we see how these two populations evolve relative to one another.

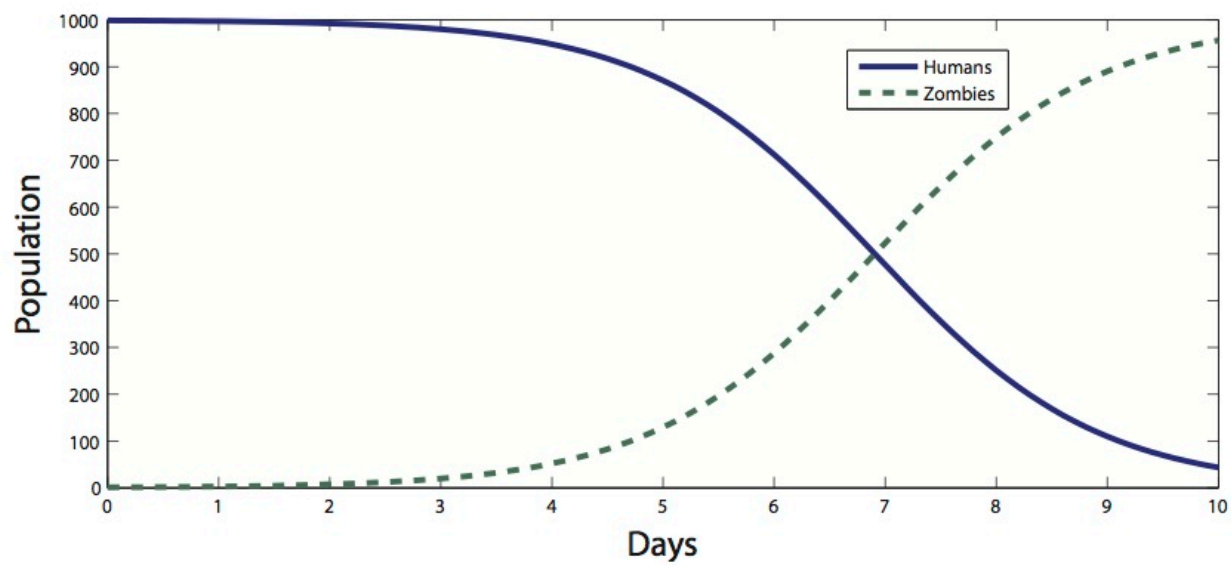
S: humans
β: transmissibility

Who will win in the human-zombie war?



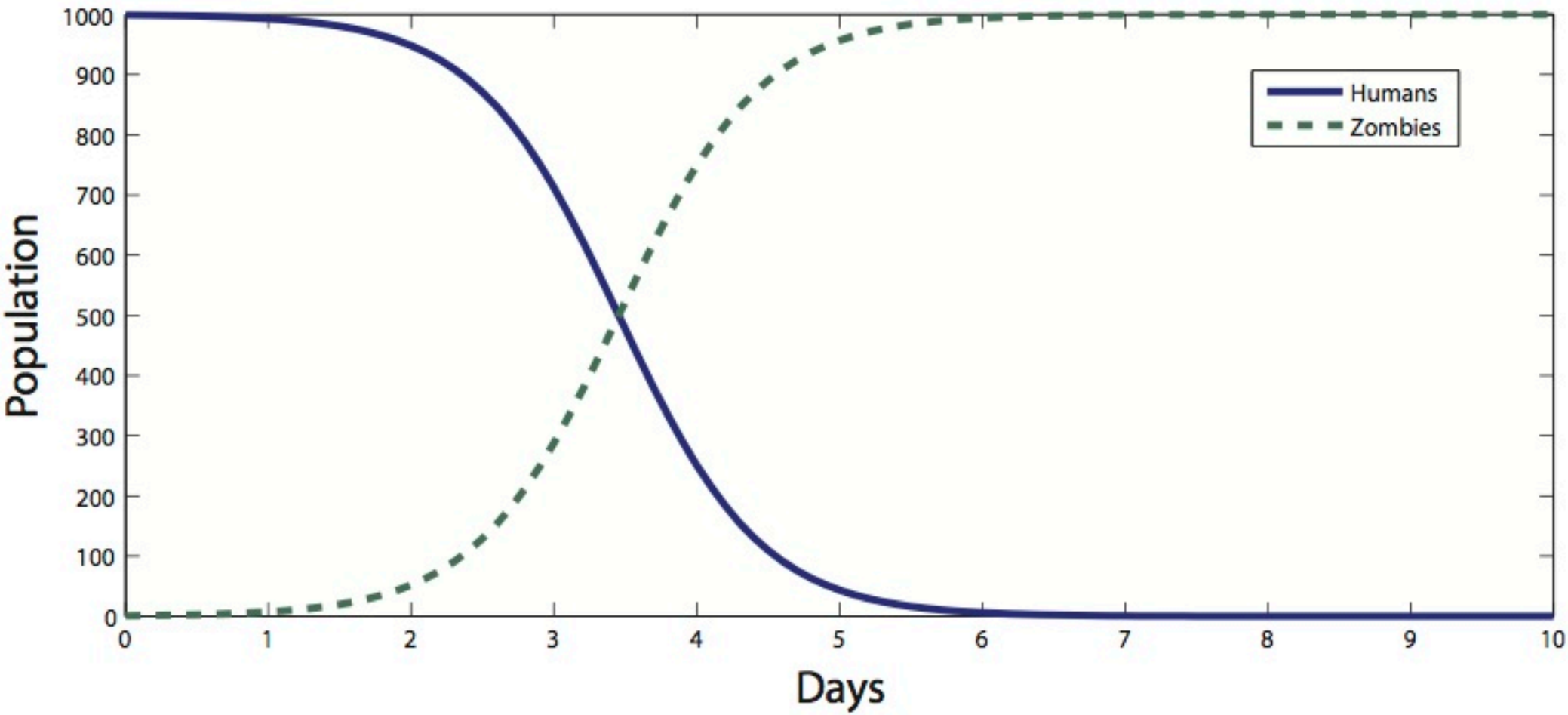
The ten day war

- Thus, a population of 1000 humans diminishes to less than 10% after 10 days
- Based on a single zombie initially attacking and converting humans at a rate of one human per zombie per day
- This was for β=0.001
- What if we doubled β?



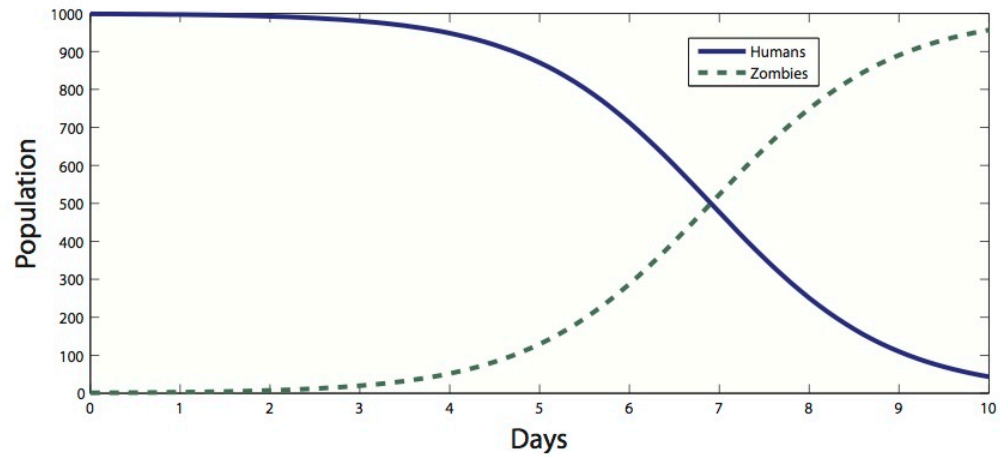
β: transmissibility

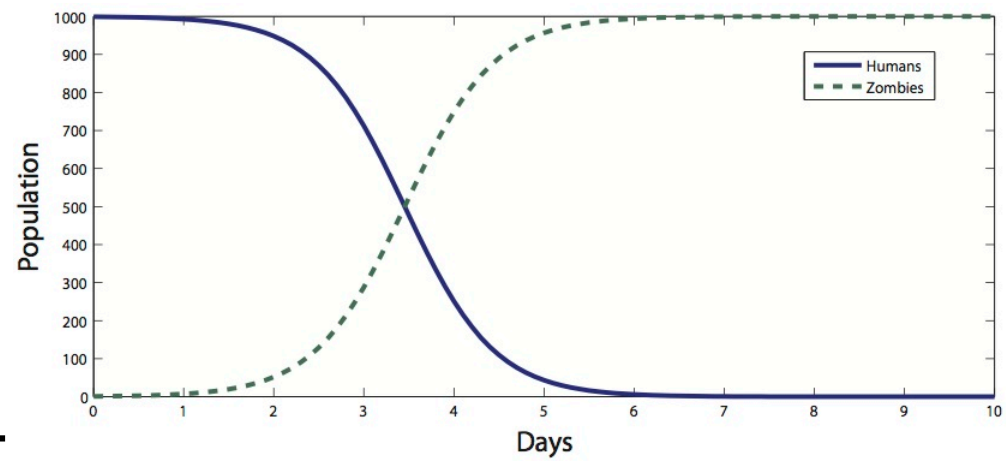
A faster infection rate



Quantitative differences

- In this case, the population dwindles to less than 10% after just 5 days
- By the seventh day, the humans have effectively died out
- The qualitative behaviour is the same, but the quantitative implications are different.





A reversal

- These predictions were made under the assumption that we know the exact rate at which zombies attack and convert humans
- More realistically, we would not know β, but instead have (inaccurate) observations of how the population changes over time
- The goal in this case is then reversed:
- Given some observations regarding the population at certain time points, we want to estimate β such that our model best describes the situation as we see it.

β: transmissibility

Inverse problem

- Estimating rate constants from observation is very challenging
- This is known as the *inverse problem*
- However, it does come with benefits
- Once we have inferred β, we may then quantify the likelihood of future scenarios, based on the knowledge that our model adequately describes the past.

β: transmissibility

A caveat

- The model is based on simplifying assumptions
- It does not capture every detail of the physical system
- There will also be measurement errors and uncertainty in the data
- In a zombie attack, humans in hiding may go unrecorded and zombies may lurk unnoticed in dark corners
- Thus, there is no single “true” value of β.

β: transmissibility

Most likely values

- There may be many values of β that are approximately as good as each other at describing the data
- It makes sense then to determine the probability distribution over the *most likely* values for β
- As opposed to a single "best" value
- Statistical inference, particularly Bayes' theorem, gives us a mathematical framework in which to carry out these calculations in terms of probability distributions.

β: transmissibility

The first level of inference

- Determining parameters with which the model plausibly describes the data
- This is the probability of the free parameters $\theta=[\theta_1,..., \theta_n]$ given some data Y and a particular model M
- We write this as $P(\theta|Y,M)$
- In our example:
  - $\theta=\beta$
  - Y is a vector of observations at a number of time points.

β: transmissibility

The second and third levels of inference

- The second sheds light on the uncertainty associated with the choice of model
- The probability of a particular model M given the data Y
- We write this as P(M|Y)
- The third describes the probability of a *prediction*, given the data
- This prediction may be based on multiple plausible models which are weighted according to their relative probabilities.

M: particular model
Y: vector of observations

Bayes' theorem

- In order to estimate these probability distributions, we can use Bayes' theorem
- This gives us a method of combining prior knowledge and newly obtained data
- Key to the Bayesian approach is the idea that observations are inherently uncertain
- Thus, a single data point is assumed to be just one sample from some underlying probability distribution.

The prior

- This uncertainty is represented as the likelihood of the data given a model and its current set of parameters
- This is written as $P(Y|\theta,M)$
- The prior distribution $P(\theta|M)$ characterises our initial knowledge or belief regarding plausible values of the parameters
- This is often simply referred to as “the prior”.

M: particular model
θ: free parameters
Y: vector of observations

Applying this to our zombie outbreak

- A prior can be constructed by considering a reasonable timescale for the process
- eg all the action will not be over in one day
- Thus β has an upper bound of 1
- With this value, $S(0)\times\beta=S(0)$
  - ie all S(0) humans could be converted by one zombie on the very first day
- Likewise, a lower bound on β is zero
  - in this case, zombies never convert humans

S: humans
β: transmissibility

Uniform distribution

- In the absence of further information, we could decide that any value of β between 0 and 1 is equally likely
- So we may choose our prior on β to have a uniform distribution over this range
- Thus, we already have some (limited) information on our key rate constant before we gather any data.

β: transmissibility

The likelihood

- A measure of the goodness of fit between the data and the output of the model
- The choice of which probability distribution to employ depends on the problem context
- A Poisson distribution may be appropriate
  - eg if the observed data is the number of counts occurring within a particular time interval.

Gaussian distribution?

- Alternatively, the observed data may be obtained from estimates which may be affected by a large number of small but unknown random factors
- Then, due to the Central Limit Theorem, the associated error may be well approximated by a Gaussian distribution.

Independent errors

- For our zombie attack, we assume that the estimated population levels are subject to small, unknown errors
- The final estimates will combine local intelligence, large numbers of individual sightings etc
- We also assume that errors at different observation times are independent.

The likelihood function

- We define it to be a quantitative measure of the agreement between the model output and the observed data over all time points:

$$L = P(\mathbf{Y}|\boldsymbol{\theta}, M) = \prod_t N_{Y(t)}(S(t), \sigma^2)$$

  - P(Y|θ,M) represents the probability of an observation given a model M with parameters θ
  - Y(t) is the observation at time t
  - S(t) is the output of the model at time t, given parameters θ; and

$$N_x(\mu, \sigma^2) \equiv \frac{\exp(-(x-\mu)^2/(2\sigma^2))}{\sqrt{2\pi\sigma^2}}$$

    is the density for a Gaussian with mean μ and variance σ^2.

M: particular model
θ: free parameters
Y: vector of observations
S: humans

Variance

- The variance σ^2 is the inherent level of uncertainty in the data
- It could be estimated and fixed in advance
- Or inferred along with other parameters.

In summary

- Thus, for a particular combination of model parameters and initial conditions, given an observation of the number of surviving humans at a known point in time, we compute the likelihood by taking a Gaussian density centred on the model prediction
- We then find the value of the density function at the observed value
- We repeat this for each data point and multiple the answers together.

Updating our initial beliefs

- Bayes' theorem allows us to update our initial belief about the parameter values, as defined by the prior, by taking the data into account
- Our updated knowledge is then quantified by the posterior distribution P(θ|Y,M) by combining the prior distribution with the likelihood function:

$$P(\boldsymbol{\theta}|\mathbf{Y}, M) = \frac{P(\mathbf{Y}|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M)}{P(\mathbf{Y}|M)}$$

$$\propto P(\mathbf{Y}|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M).$$

M: particular model
θ: free parameters
Y: vector of observations

Marginal likelihood

- The marginal likelihood P(Y|M) is constant for a particular model M
- Thus, it may be calculated as the integral of the likelihood times the prior over all parameter values:

$$P(\mathbf{Y}|M) = \int \ldots \int P(\mathbf{Y}, \boldsymbol{\theta}|M) d\theta_1 \ldots \theta_n$$

$$= \int P(\mathbf{Y}|\boldsymbol{\theta}, M) P(\boldsymbol{\theta}|M) d\boldsymbol{\theta}.$$

M: particular model
θ: free parameters
Y: vector of observations

The challenge of inference

$$P(\mathbf{Y}|M) = \int \dots \int P(\mathbf{Y}, \boldsymbol{\theta}|M) d\theta_1 \dots \theta_n$$
$$= \int P(\mathbf{Y}|\boldsymbol{\theta}, M) P(\boldsymbol{\theta}|M) d\boldsymbol{\theta}.$$

- This integral is generally analytically intractable and high dimensional
- This makes the second and third levels of statistical inference over ODE models challenging
- Recently though it has been shown that this integral may be efficiently and accurately estimated using a technique called thermodynamic integration
- We'll come back to this.

M: particular model
θ: free parameters
Y: vector of observations

Metropolis-Hastings

- Finally, we need a method of sampling from the posterior distribution
- In our case, this is analytically intractable
- We can't calculate the marginal likelihood for our models based on nonlinear ODEs
- Instead, we can employ the Metropolis-Hastings algorithm:
  - a Markov chain Monte Carlo method
  - draws a series of random samples from an approximation of the posterior
  - feasible thanks to the power of modern computing.

Sampling

- In Bayesian statistics, Metropolis-Hastings allows us to draw accurate samples from the posterior distribution even if the marginal likelihood P(Y|M) is not known
- Let β_c be a current value of our parameter
- We randomly generate a new state β_n from a *proposal distribution* $Q(\beta_n|\beta_c)$ which depends only on the current state
- This should be as similar as possible to the target distribution you wish to sample from.

M: particular model
θ: free parameters
Y: vector of observations
β: transmissibility

Determining a new state

- We have little advance information about the posterior distribution
- It suffices to employ a Gaussian distribution with mean β_c and some variance chosen to give an acceptance rate of 20-40%
- This new state is accepted with probability

$$\min\left\{\frac{P(\beta_n)Q(\beta_c|\beta_n)}{P(\beta_c)Q(\beta_n|\beta_c)}, 1\right\}$$

  where P(β) is the probability distribution we wish to sample from
- This is the posterior P(β|Y,M).

M: particular model
Y: vector of observations
Q: proposal distribution
β: transmissibility

Hot spots

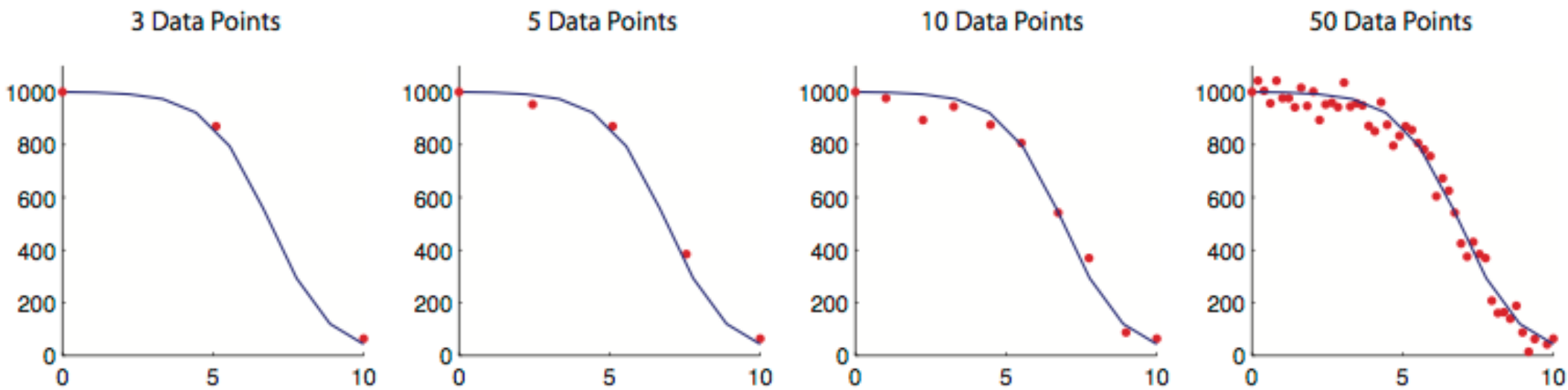
- This technique allows us to search through the set of possible parameter values
- We spend most of our time in the “hot spots” where parameter values are most promising
- We are now ready to infer a posterior distribution over the parameter values, given some data
- We’ll do this using “artificial” data
- This way, we can judge the performance of the algorithm under controlled circumstances.

Generating data sets

- We'll evaluate the solution of the ODEs for a chosen value of β at a number of time points
- We'll then add some Gaussian-distributed noise with known variance to the solution
- This will generate some experimental data
- We generated four data sets this way.

β: transmissibility

Four data sets

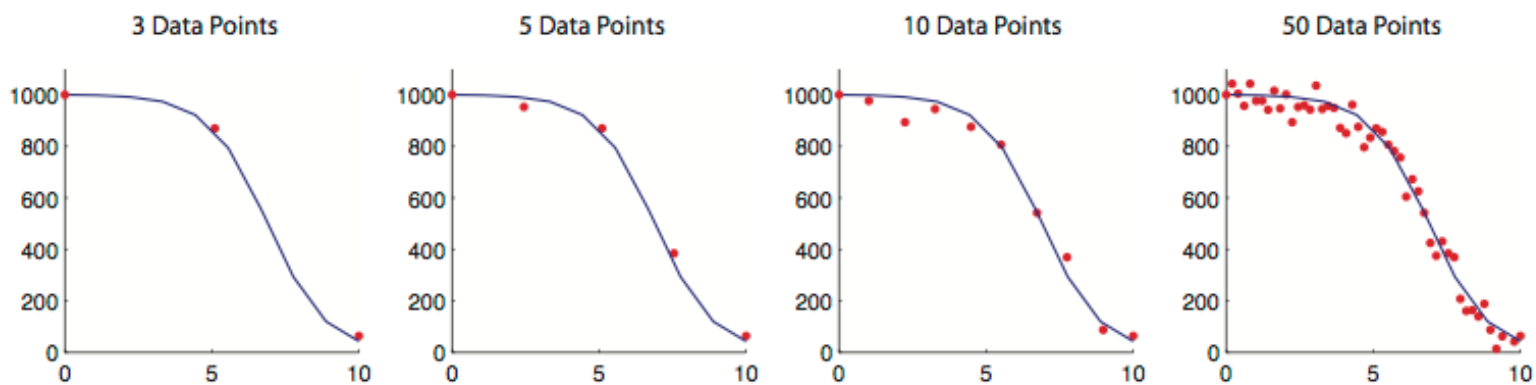


- Data sets generated from the simple zombie model over 10 days
- β=0.001
- Gaussian-distributed noise with a standard deviation of 50.

β: transmissibility

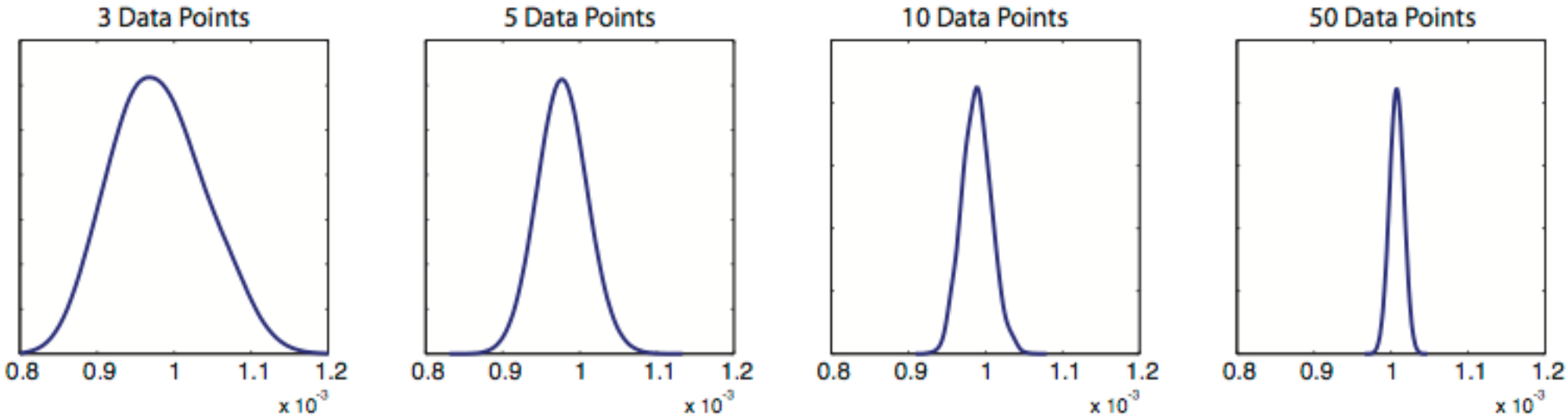
Quality of inference

- We then treat this data as though it came from the model with an unknown value of β
- Now we want to see what quality of inference is possible.



β: transmissibility

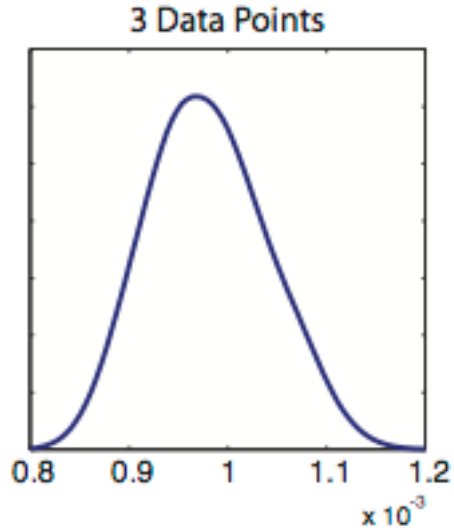
Posterior peaks



- Posterior output from the simple zombie model
- β=0.001 and ten data points
- As the standard deviation of the added noise decreases, the posterior becomes more sharply peaked around the true value of β.

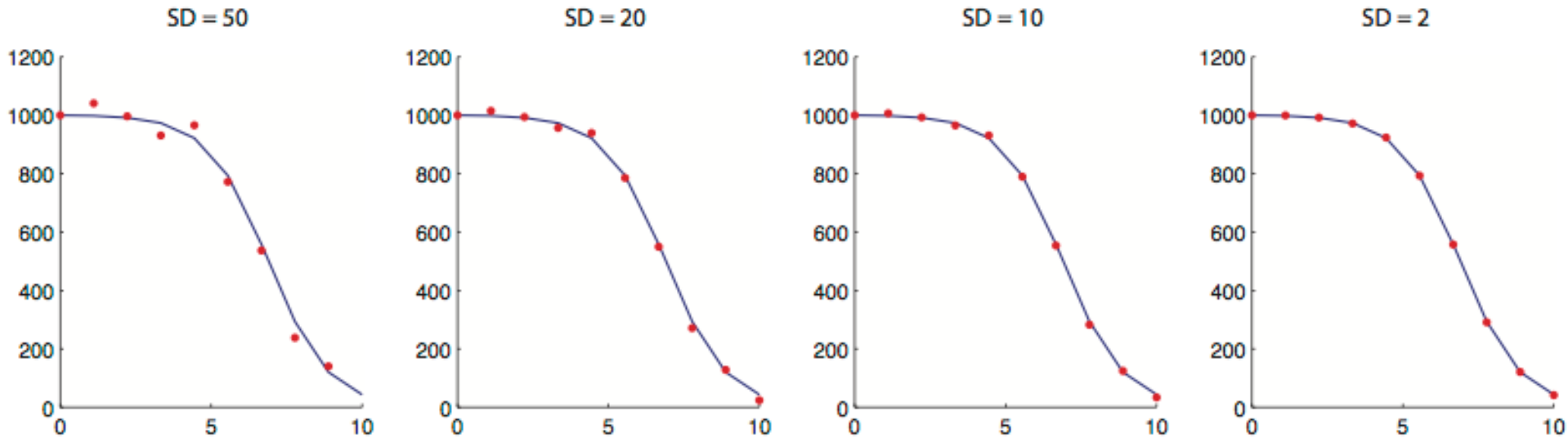
β: transmissibility

Noise-induced bias



- Noise induces a noticeable bias when using just 3 data points
  - although the posterior is reasonably large at the true value of 0.001
- However, if we add less noise, the posterior distribution becomes less diffuse
- This indicates a greater confidence in the range of values for which the model could plausibly describe the data.

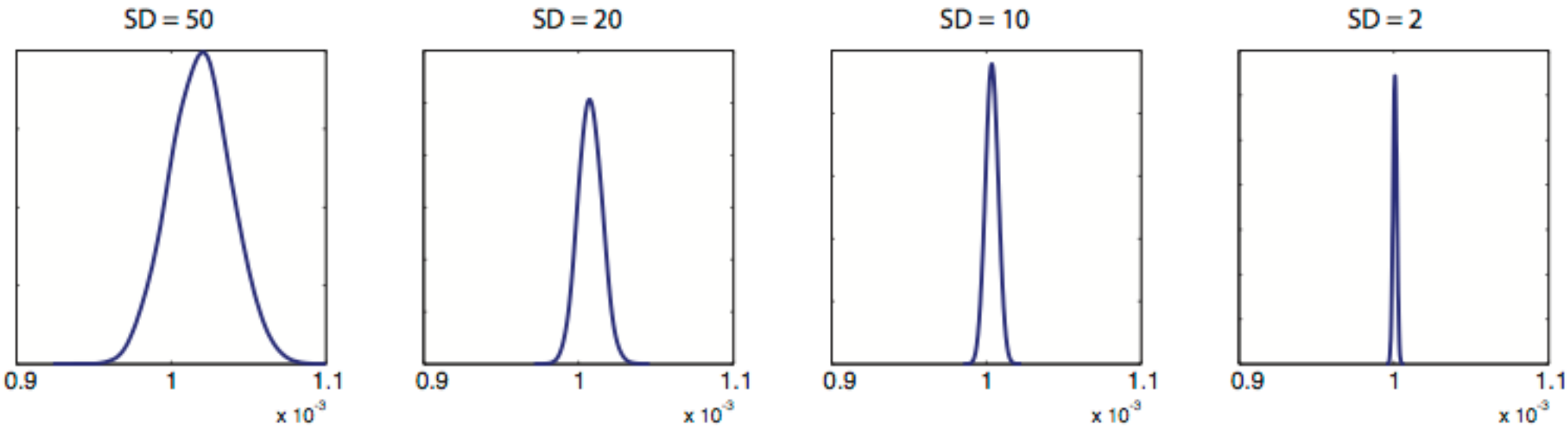
Decreasing noise



- Ten data points generated from the zombie model
- β=0.001
- Gaussian-distributed noise with standard deviation of 50, 20, 10 and 2.

β: transmissibility

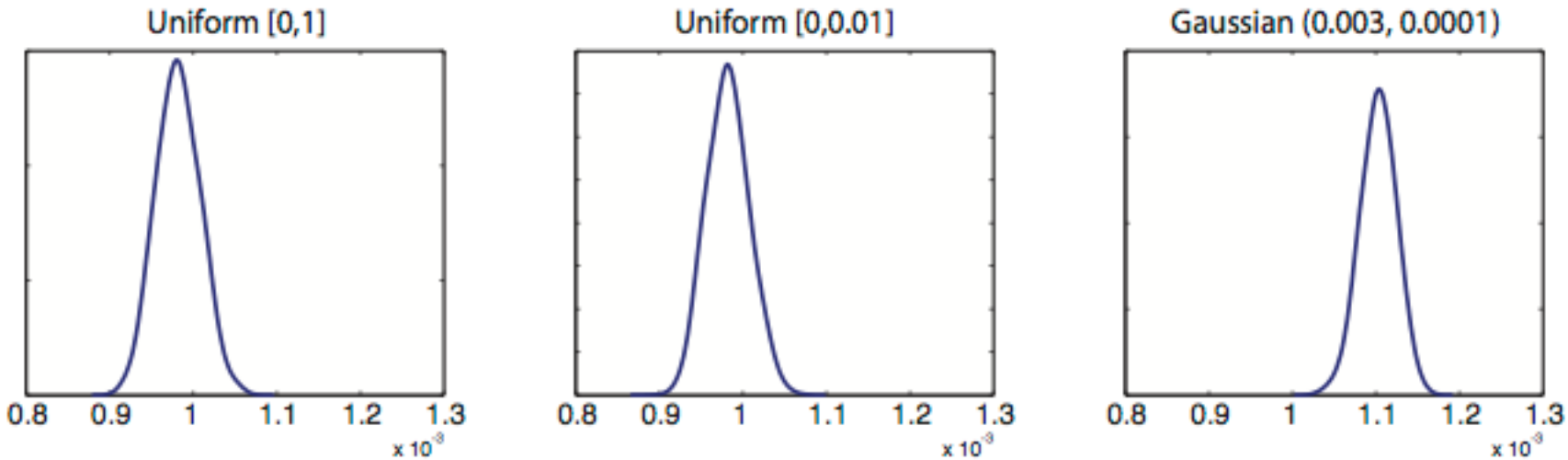
Sharper peaks



- Posterior output from the simple zombie model
- As the standard deviation of the added noise decreases, the posterior becomes more sharply peaked around the true value of β.

β: transmissibility

What if we change the prior?



- Changing the prior to uniform over [0,0.01] has little effect on the posterior of β
- If we badly mis-specified the prior, we observe a biased and skewed posterior distribution
  - this type of mis-specified prior can be diagnosed by comparing the prior and posterior.

β: transmissibility

More realistic model

- We've thus shown how Bayesian inference can be applied in a very simple ODE setting
- We next move to a more realistic model where:
  - there is more than one unknown parameter
  - the ODE solution cannot be written down explicitly
- In particular we now allow for the possibility that a human can survive a zombie attack
- In this case, the human is unscathed and the zombie joins the *removed* class.

The removed class

- The zombies are removed, rather than dead
- They can later resurrect and join the ranks of the undead
- The kill rate is α (humans and zombies fight)
- The resurrection rate is ζ
- Thus, the more realistic model is

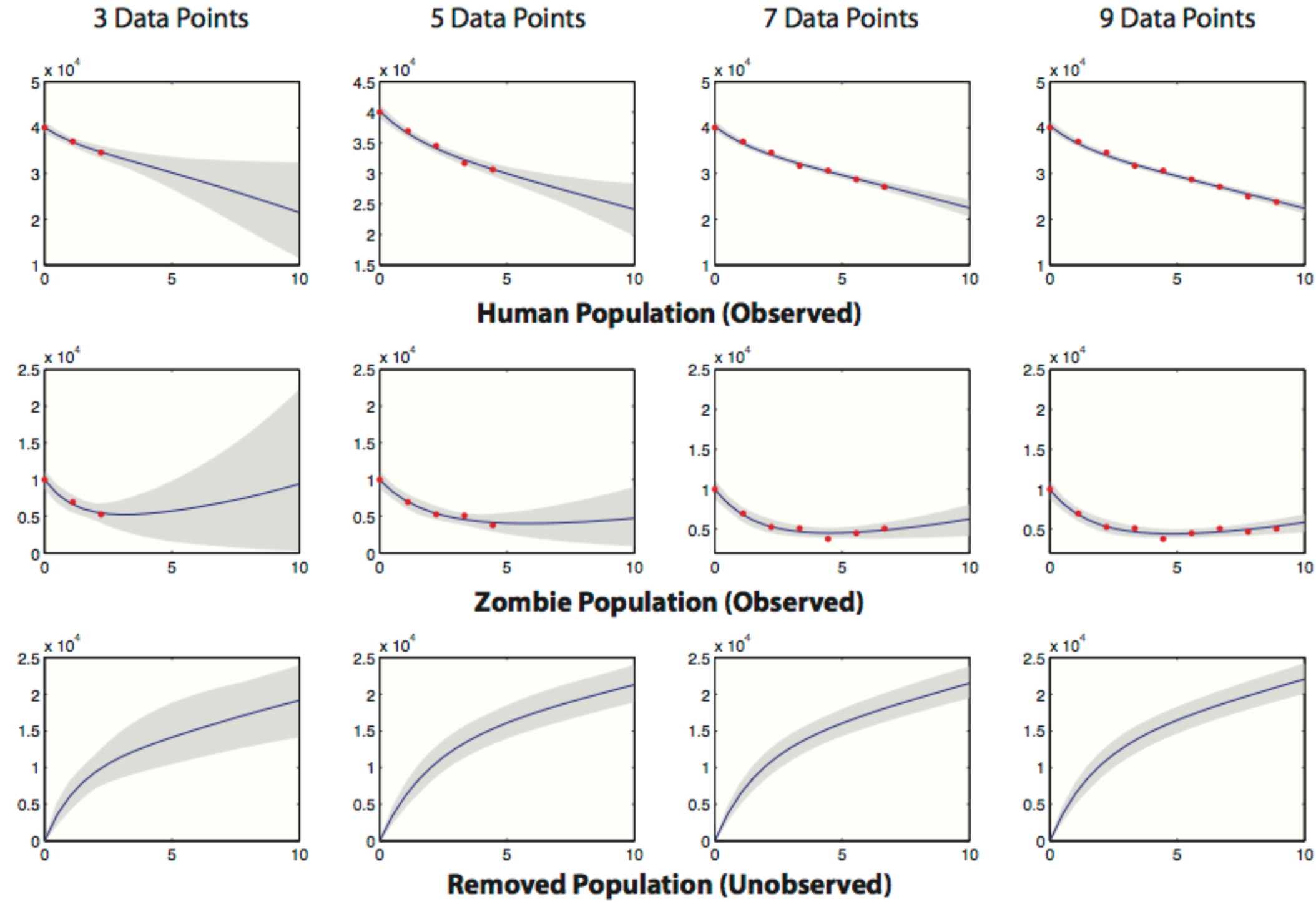
$$S' = -\beta SZ$$
$$Z' = \beta SZ + \zeta R - \alpha SZ$$
$$R' = \alpha SZ - \zeta R$$

- This is the model from Munz et al.

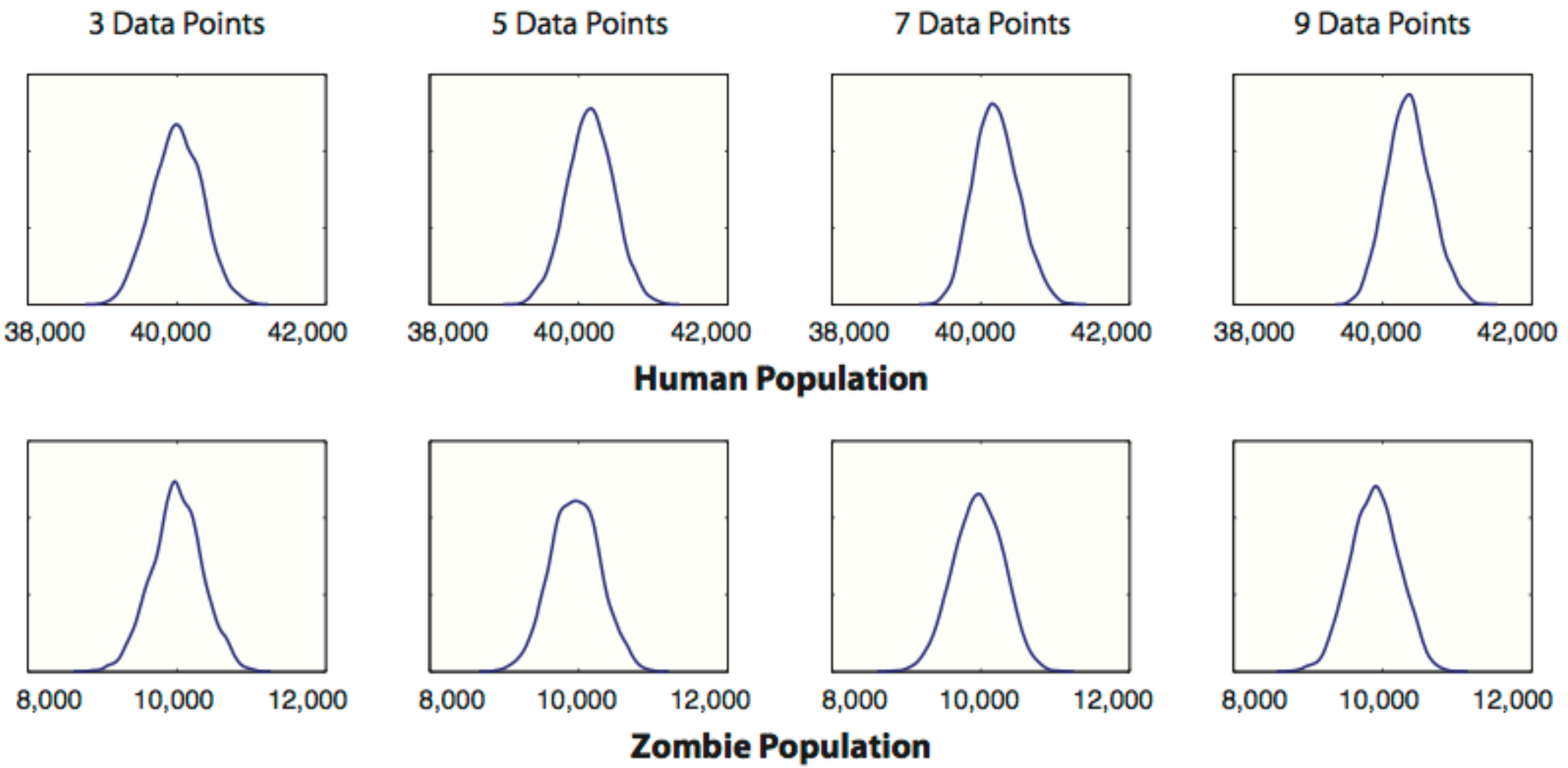
S: humans Z: zombies
β: transmissibility

Attacking a small town

- Suppose zombies attack a small town
  - eg Wakefield
  - 40,000 humans living in the town
  - 10,000 zombies attacking from Ottawa
- We take daily observations over a period of 3, 5, 7 and 9 days
- We want the *predictive model output*
- ie two standard errors or 95% confidence for the output of the ODE at each time point after the first zombie attack.


3 Data Points
5 Data Points
7 Data Points
9 Data Points
Human Population (Observed)
Zombie Population (Observed)
Removed Population (Unobserved)

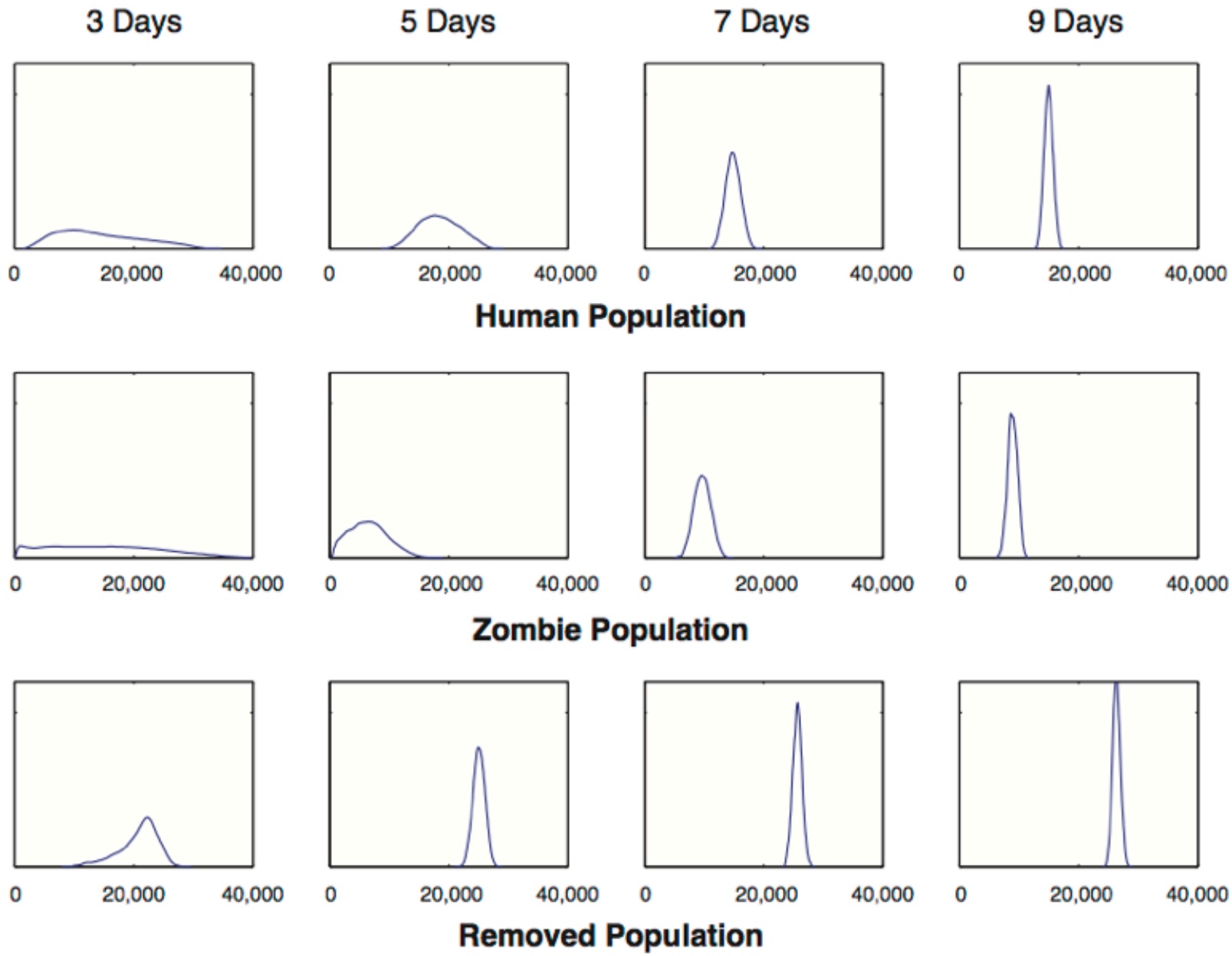
Posterior distributions of inferred initial conditions



Decreasing uncertainty

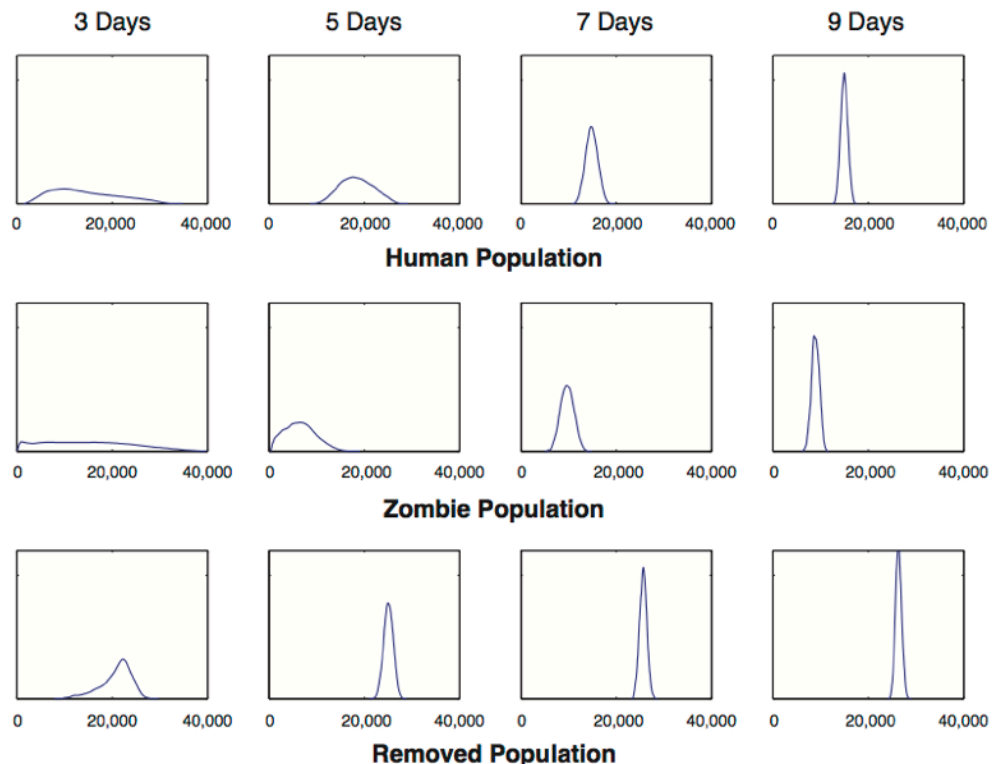
- The initial conditions are relatively insensitive to the number of data points observed
- As the number of data points increases, the uncertainty in the predictive model decreases
- Consider the predictive posterior model output for day 15, given observations over 3, 5, 7 and 9 days.

Predicted levels on day 15



Increasing confidence

- With data for just 3 days, we learn very little
- Given additional observations, we can predict with much greater certainty that the number of surviving humans is between 10,000 and 25,000
- As we collect more data, our predictions become more confident.



“True” values

- The predicted ranges are tending towards the “true” number of humans
- Determined by the system of ODEs to be 14,790
- Likewise, the predicted number of zombies tends towards the “true” value of 10,426
- Removed individuals tends towards 24,784.

Model selection

- We now consider a second level of inference
- With uncertainty not only in the parameters, but also in the specified model
- Suppose during the attack on Wakefield that there are rumours that the zombies will only attack in pairs
- To simplify things, we can assume that a single zombie who meets a human will always try to flee
- Similarly with a single human who meets a pair of zombies.

The pair-attack model

- When a single zombie encounters a susceptible, either both remain unscathed or the zombie becomes removed
- When a pair of zombies encounters a susceptible, either all remain unscathed or the susceptible succumbs to zombification
- Thus, the pair-attack model is

$$S' = -\beta S Z^2$$

$$Z' = \beta S Z^2 + \zeta R - \alpha S Z$$

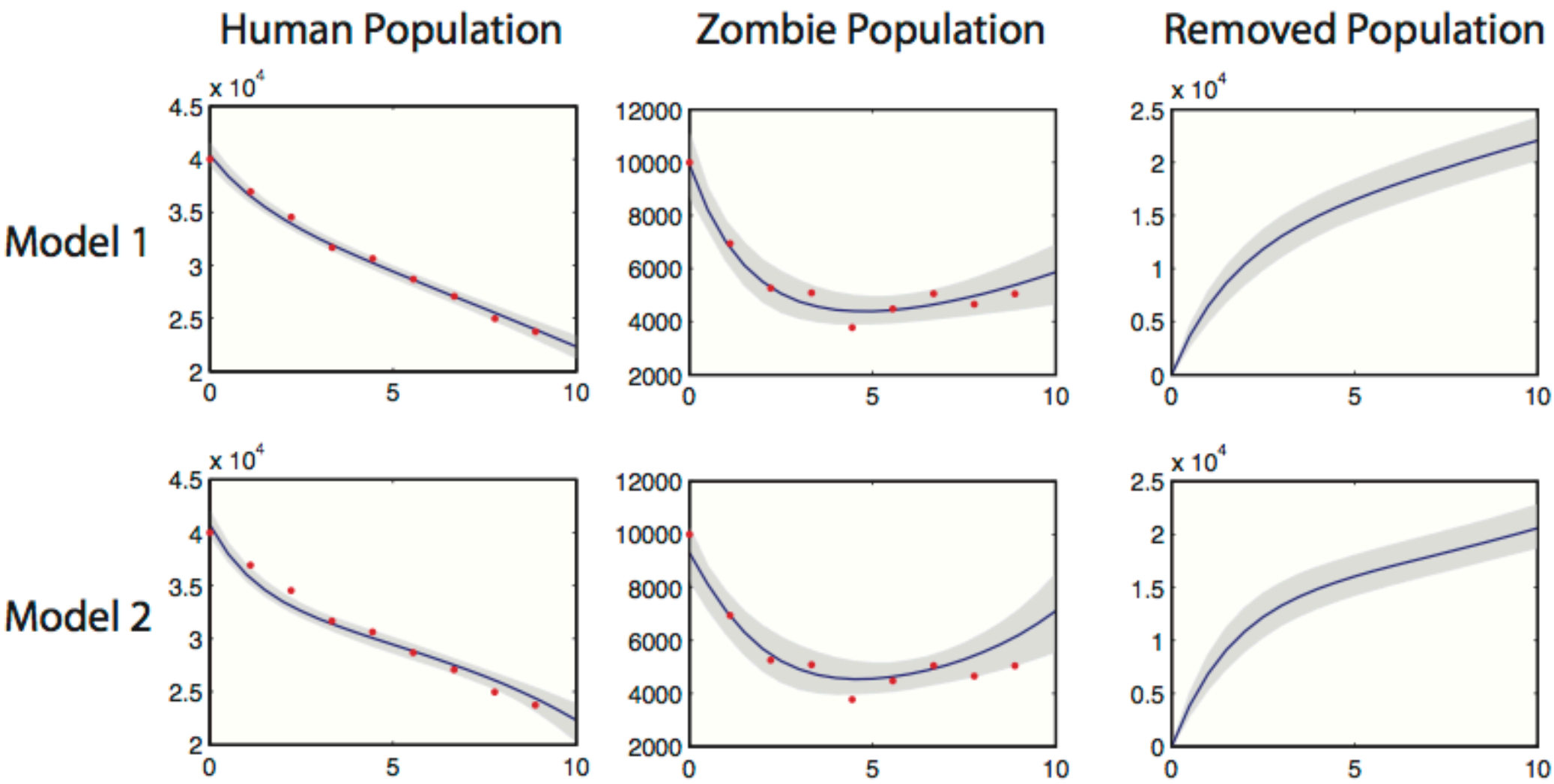
$$R' = \alpha S Z - \zeta R.$$

S: humans Z: zombies
R: removed
β: transmissibility
α: attack rate
ζ: resurrection rate

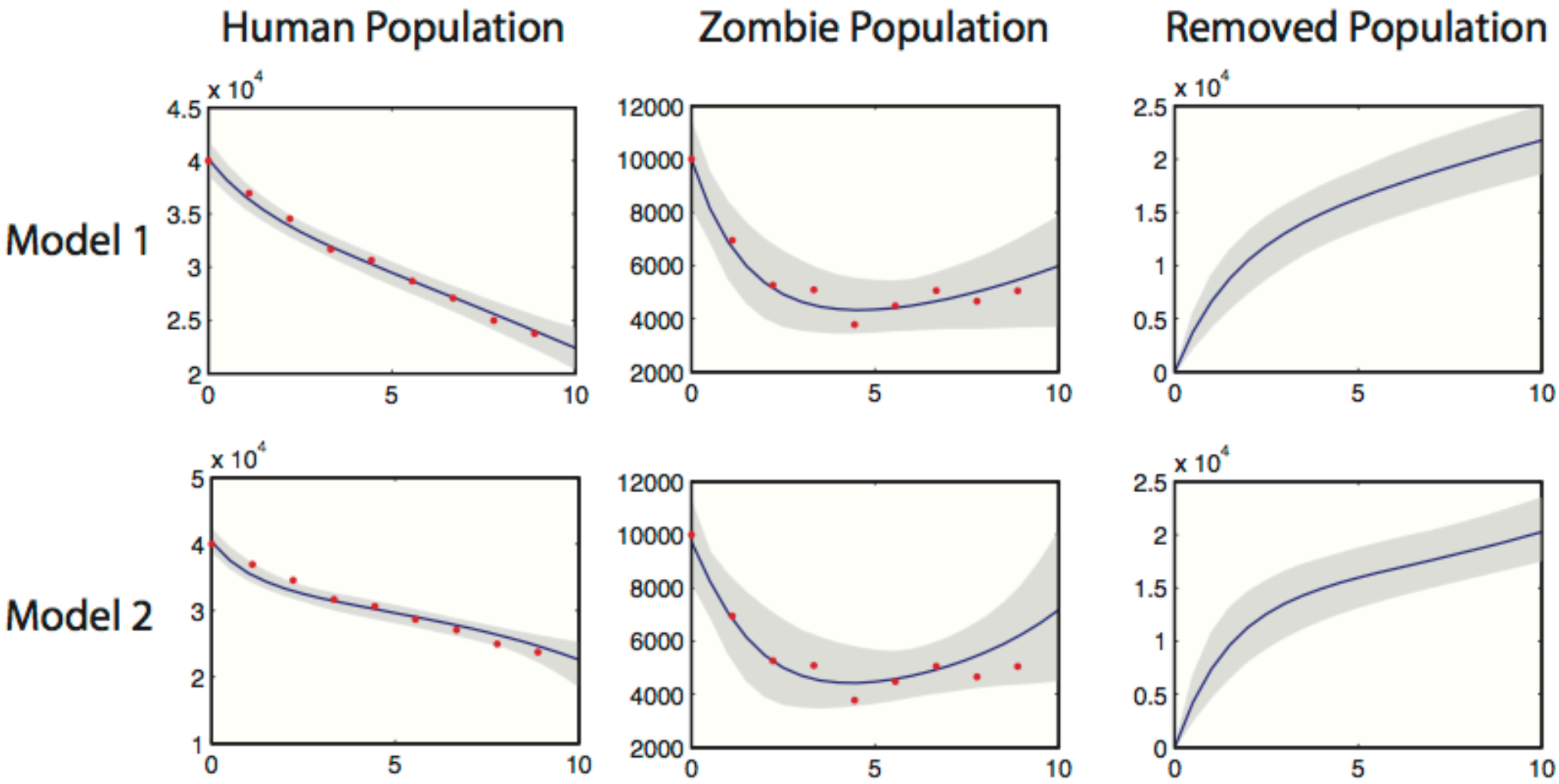
Generating observations

- We can generate observations over nine days by
  - simulating from Model 1
  - adding Gaussian-distributed noise with standard deviation 500, 1000 or 2000
- We have no information about the parameters
- Will an inference algorithm allow us to conclude that Model 1 describes the data better than Model 2?

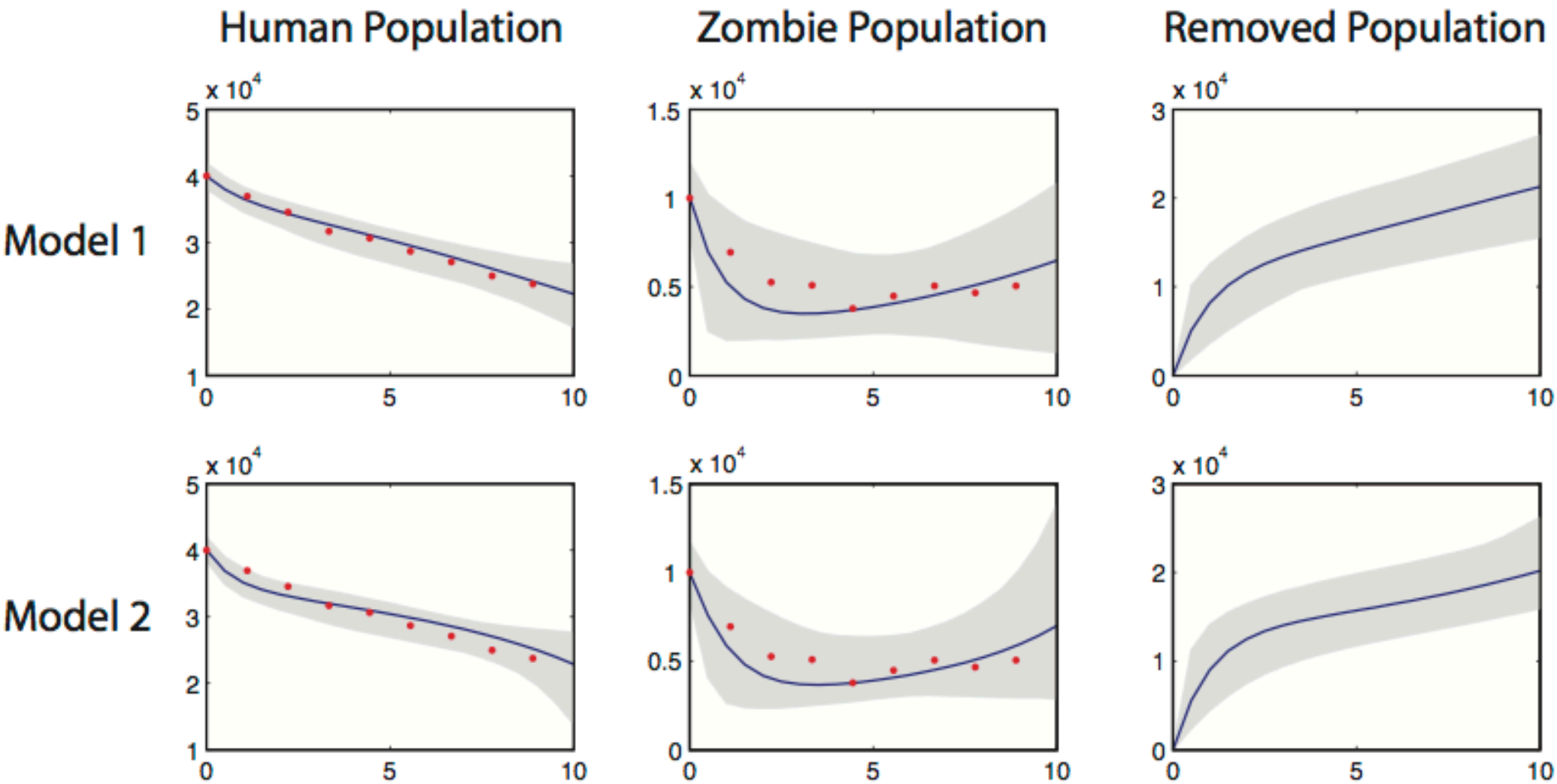
Posterior output, SD=500



Posterior output, SD=1000



Posterior output, SD=2000



Bayes factors

- Visually assessing which is the better model is difficult
- The posterior output covers most of the data points for both models
- Instead, we can calculate *Bayes factors*
- B_{12} represents the weight of statistical evidence in favour of Model 1 over Model 2
- Computed as the ratio of the marginal likelihoods for the two competing models

$$B_{12} = \frac{P(\mathbf{Y}|M_1)}{P(\mathbf{Y}|M_2)}.$$

M_j: particular model
Y: vector of observations

Thermodynamic integration

- Calculating the marginal likelihood involves estimating the integral of the likelihood times the prior over all values of the parameters
- This is an extremely challenging task
- This is where we employ the technique of thermodynamic integration
- It has recently been shown to provide accurate, low variance estimates of this quantity
- Other, seemingly similar methods may fail to produce usable results.

Interpretation of Bayes factor

B_{12}	Evidence against alternative
1 to 3	Slight
3 to 10	Substantial
10 to 100	Strong
>100	Decisive.

B_{12}: Bayes factor

Marginal likelihoods for each model

Model	SD of Gaussian-distributed noise	Marginal log-likelihood (± Standard error)	Log(Bayes Factor) $Log(B_{12})$
Model 1 Model 2	500	-152.7 (±0.1) -184.2 (±1.7)	31.5 (±1.8)
Model 1 Model 2	1000	-158.5 (±0.1) -175.4 (±1.0)	16.9 (±1.1)
Model 1 Model 2	2000	-167.1 (±0.1) -177.0 (±3.4)	9.9 (±3.5)

Model 1, you are a winner!

- In each case, the log Bayes factors correctly identify that Model 1 was used to produce the data
- As the noise increases, the weight of evidence as indicated by the log Bayes factor decreases
- However, the evidence remains substantially in favour of the correct model.

Is it safe to go out yet?

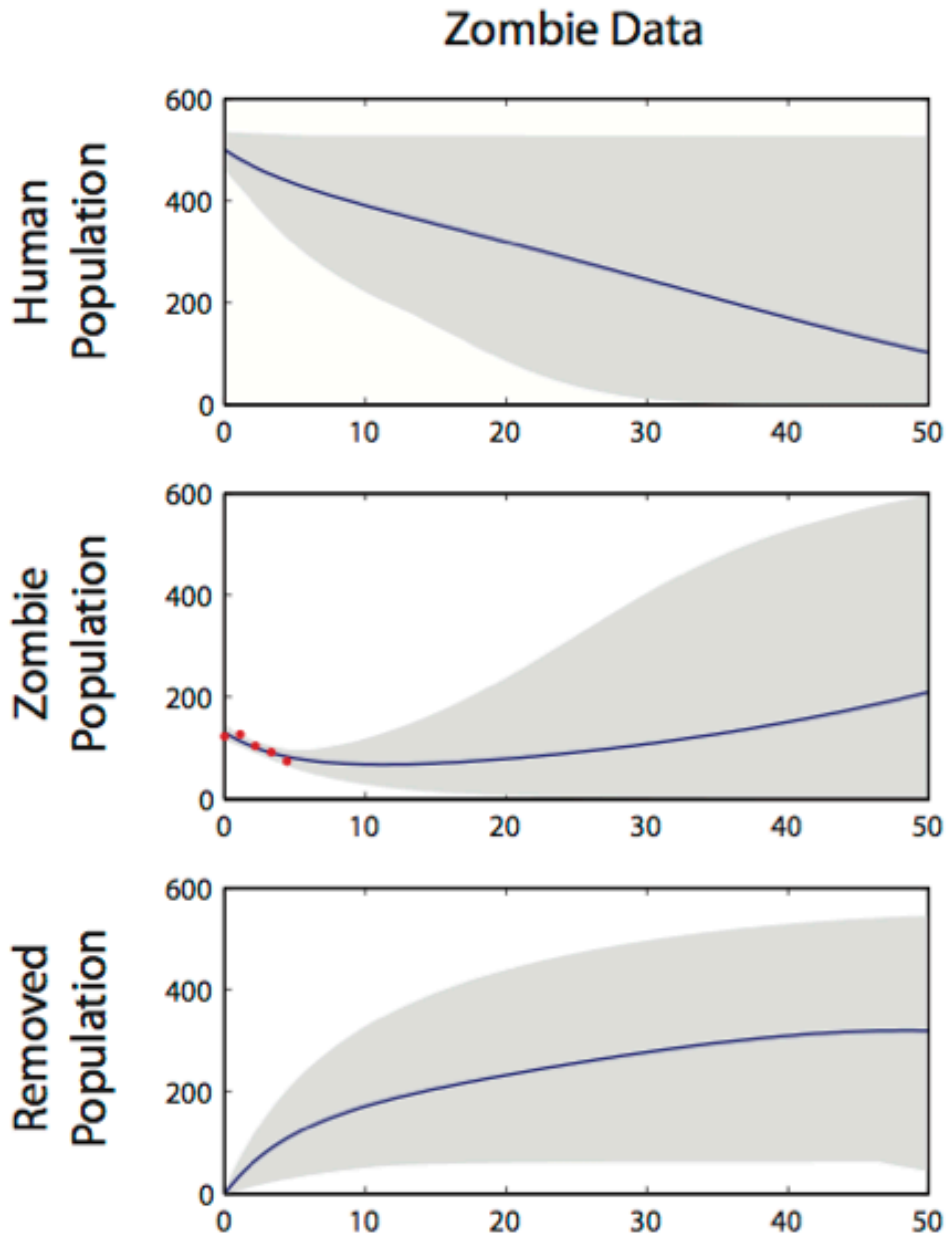
- We now return to the original question
- Soldiers patrolling the are have reported daily observed zombie numbers for the past 5 days as 123, 127, 104, 92 and 74
- We are holed up in a shopping mall with enough supplies to survive for 50 days
- Is it safe to go out yet?

Using Model 1

- We shall assume Model 1 to be a fair representation of the interaction between zombies, humans and the removed
- (If we had multiple plausible models, we could once again do a full model comparison by calculating Bayes factors)
- We perform parameter inference over the model given our data of daily observations
- Then plot the predictive model output.

Observations of zombies only

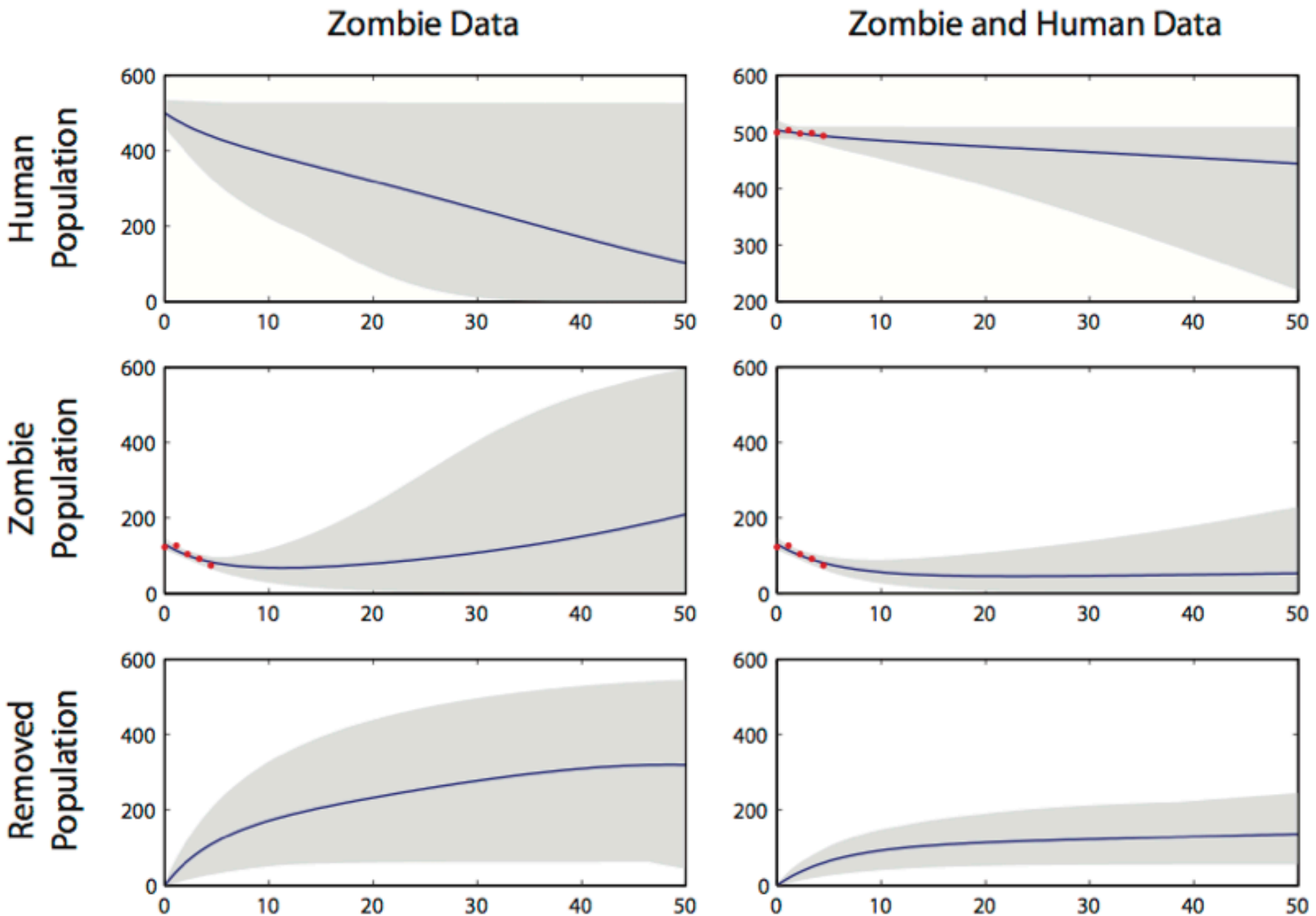
- The 95% confidence output includes a wide variety of outcomes
- Uncertainty in the estimate increases with time
- Possibilities include
  - there is relatively little impact on humans
  - zombies take over completely in a month.



Effect of uncertainty

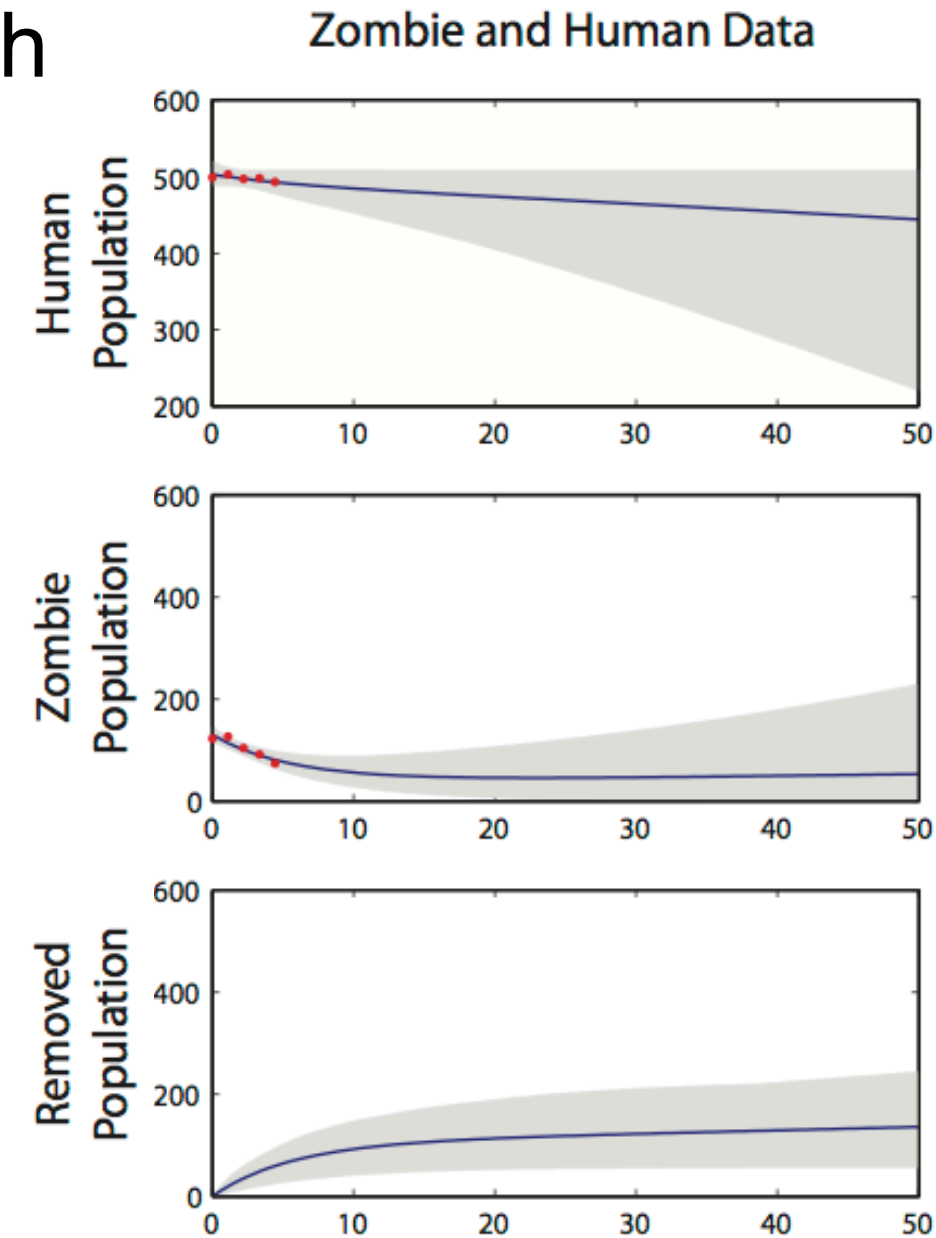
- Since the zombie population initially decreases, it might be worth making an early exit
- After that, there is much more uncertainty in the number of zombies
- What if we had more information?
- Eg the human population
- We can again perform parameter inference over our model to include data for both humans and zombies.

Predictive model output from 3 parameter model



Effect of additional information

- We can now say with much more certainty that the number of zombies will remain low for a longer period of time
- Thus, it is less urgent to escape immediately
- In this case, it's best to sit tight before gathering supplies and attempting an escape.



Summary

- Any mathematical model represents an abstracted summary that doesn't capture all characteristics of interest
- Modelling involves compromises and generates an inherent level of uncertainty
- Identifying unknown or unmeasured model parameters introduces further uncertainty
- If model predictions are used to guide policy, a systematic and consistent treatment of all levels of uncertainty is vital.

An outstanding challenge

- An outstanding challenge is the calibration of models with a large number of unknown parameters
- Requires efficient sampling in high dimensions
- Added complications arise when parameters are highly correlated.

MCMC methodologies

- It is important to solve ODEs quickly, since large numbers of solves may be required
- Highly accurate solutions are not required
- The ODE may be solved repeatedly for similar parameter values
- Another key issue is the development and analysis of new Markov Chain Monte Carlo methodologies to provide unbiased and low variance estimates of the quantities required for model comparison.

Conclusion

- Bayesian theory brings together ideas from
  - applied mathematics
  - statistics
  - computer science
- Understanding how to deal with parameters is one of the most fundamental issues in applied mathematics
- Knowing this can be the difference between safely holing up in the mall...

  ...and being eaten by a zombie because we thought it was safe to go out.

Authors

- Ben Calderhead (University of Glasglow)
- Mark Girolami (University of Glasgow)
- Des Higham (University of Strathclyde)

B. Calderhead, M. Girolami, D.J. Higham. Is It Safe To Go Out Yet? Statistical Inference in a Zombie Outbreak Model (In: R. Smith? (ed) Mathematical Modelling of Zombies, University of Ottawa Press, *in press*.)