Speeding up and extending the sensitivity analysis of an epidemiological model

John Paul Gosling

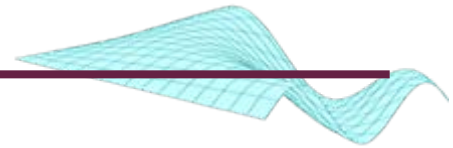
University of Sheffield

MUCM

Managing Uncertainty in Complex Models

RESEARCH COUNCILS UK
EPSRC Engineering and Physical Science Research Council
The University Of Sheffield.
Durham University
University of Southampton
LSE The London School of Economics and Political Science
Aston University

The application

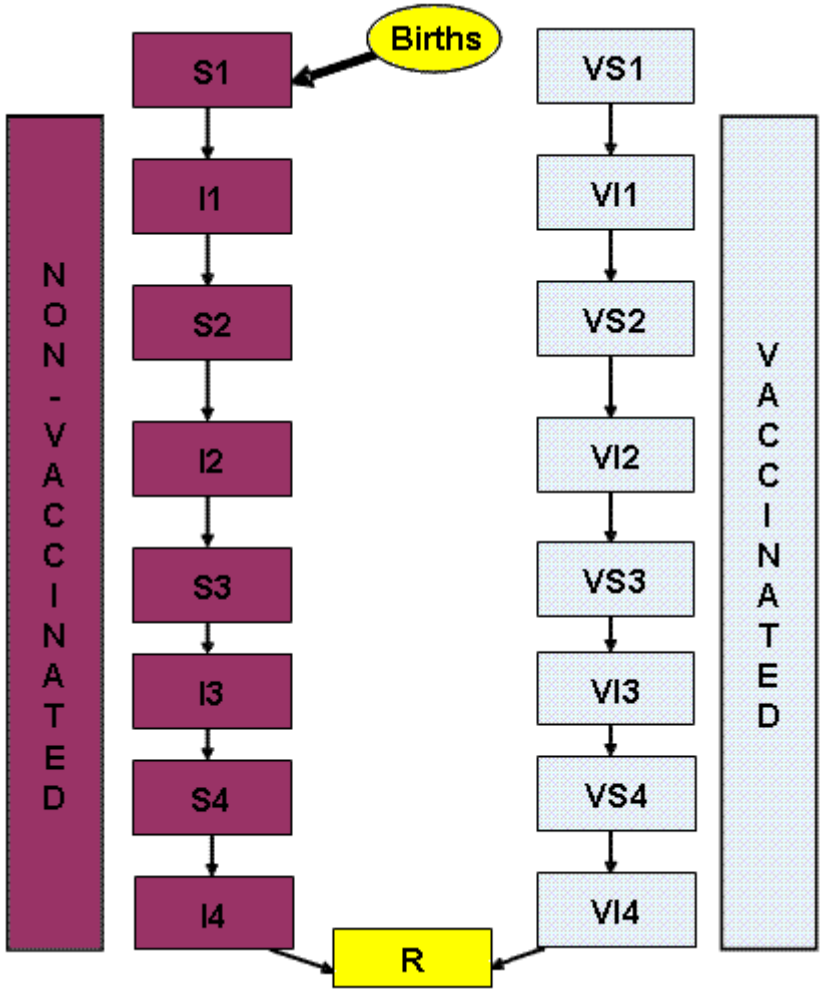


⋆ Some company is interested in a virus that is prevalent in infants.

⋆ A vaccine has been developed for this virus that could have a major impact on the third world.

⋆ The company has employed epidemiologists to model the incidence of the virus in some population and the effect of a vaccination on that incidence.

The model



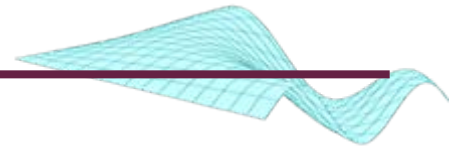
Inside the first age compartment

S = Susceptible

I = Infected

R = Recovered

The model



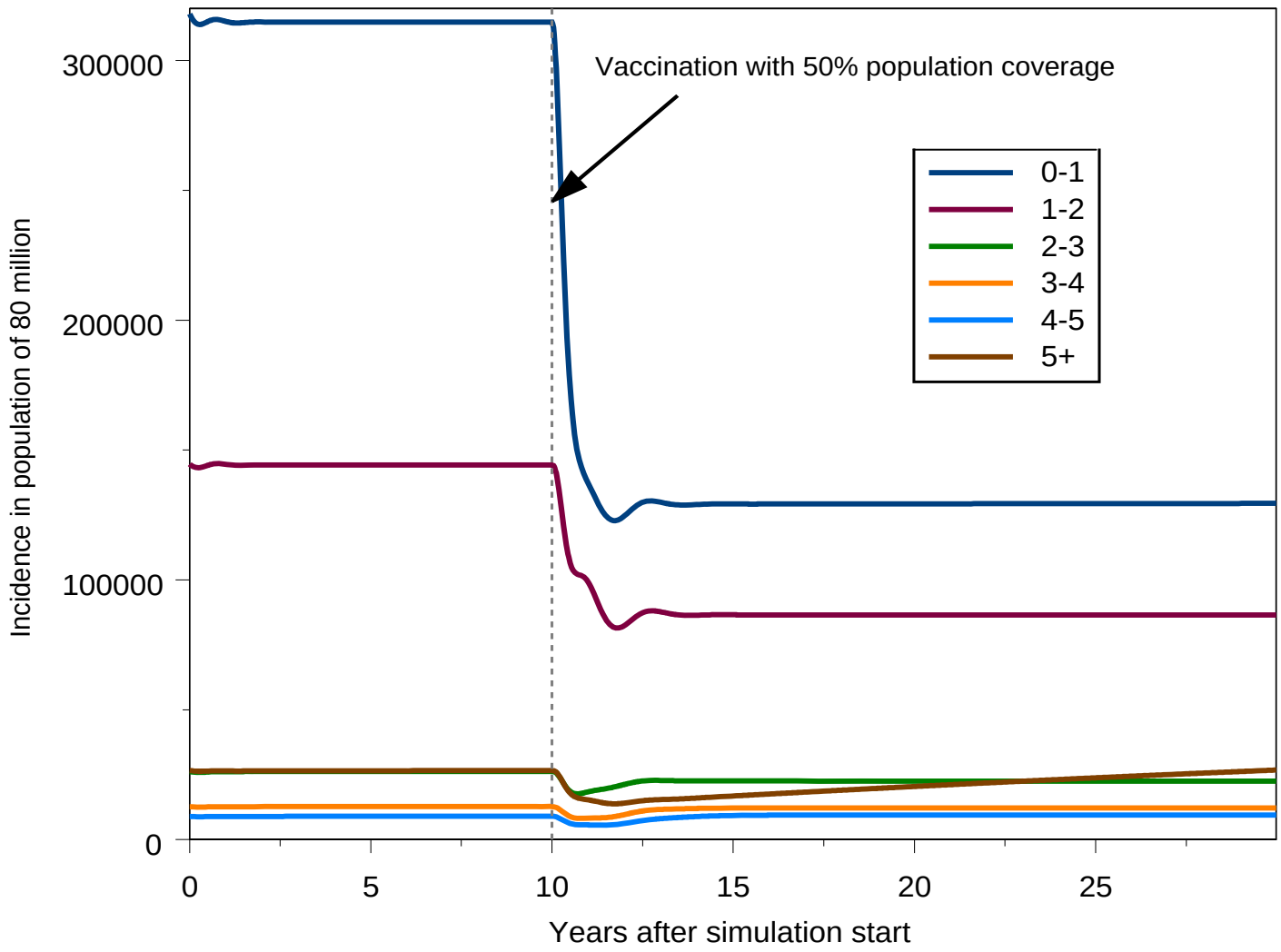
⋆ There are 672 compartments in this model.

⋆ The model takes 15 seconds to run.

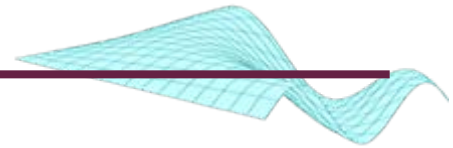
⋆ It requires 20 input parameters: 9 risk, 6 transmission, the vaccination coverage and 4 others.

⋆ The output is 6 time series of incidence for 6 age groups.

The model


Vaccination with 50% population coverage
0-1
1-2
2-3
3-4
4-5
5+
Incidence in population of 80 million
300000
200000
100000
0
0
5
10
15
20
25
Years after simulation start

Aims of the analysis

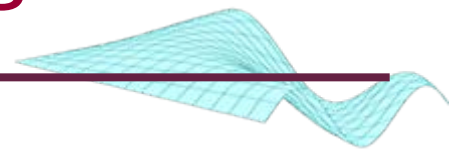


⋆ The main aim here is to help the scientists understand their model.

⋆ Previous work had involved Monte Carlo-type sensitivity analyses that cost thousands of runs. (This only included 10 of the 20 parameters).

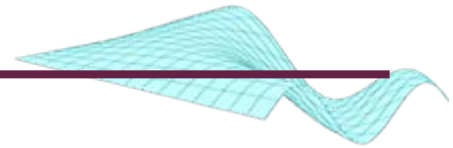
⋆ We wanted to show them that thousands of runs are not needed and we can handle all of the parameters.

An overview of the analysis



- Range elicitation
- Screening
- Formal elicitation
- Emulator building
- Emulator validation
- Sensitivity analysis

Screening

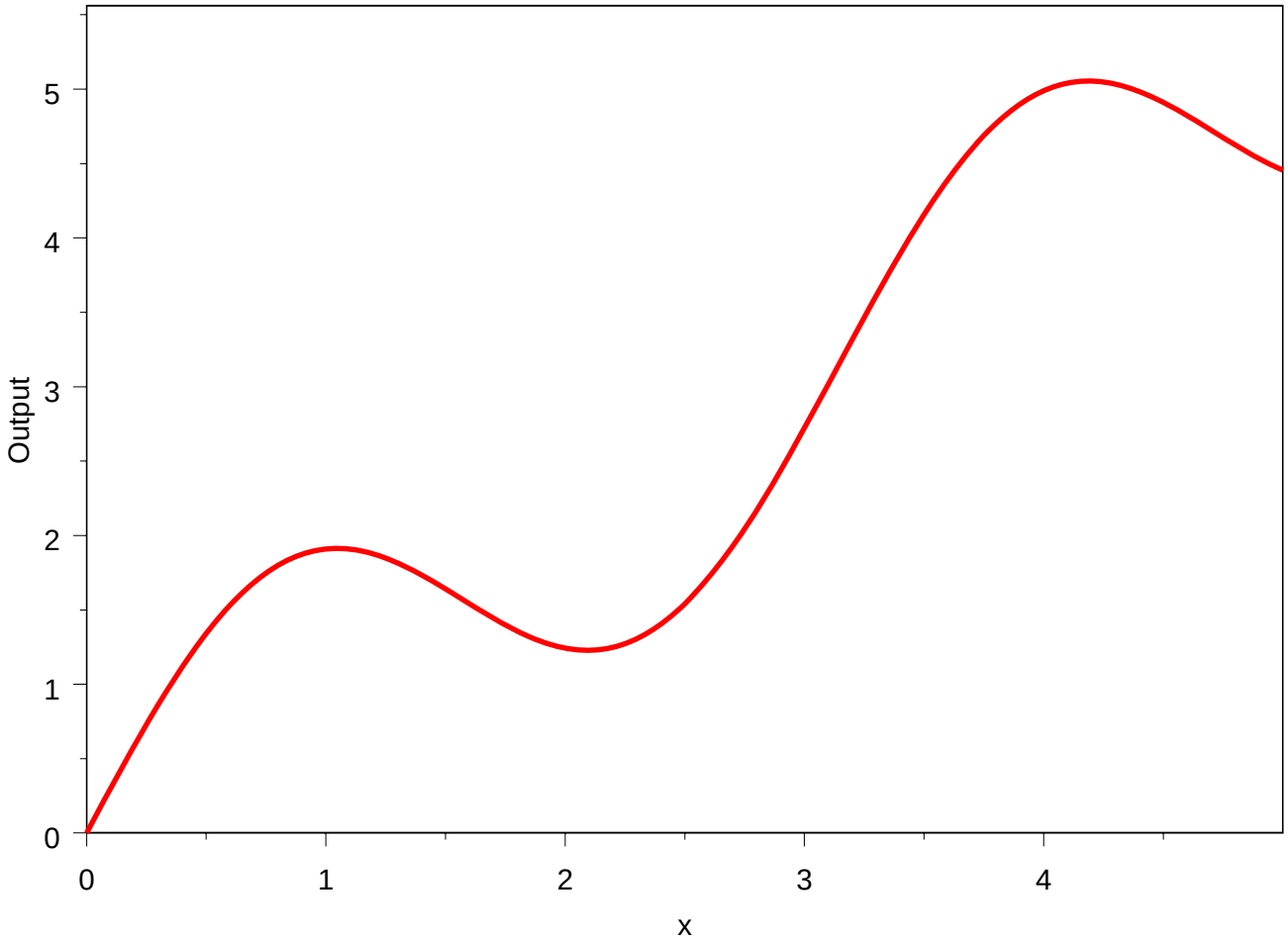


⋆ The aim of the screening stage is to identify inputs that are having no or linear effects on the output.

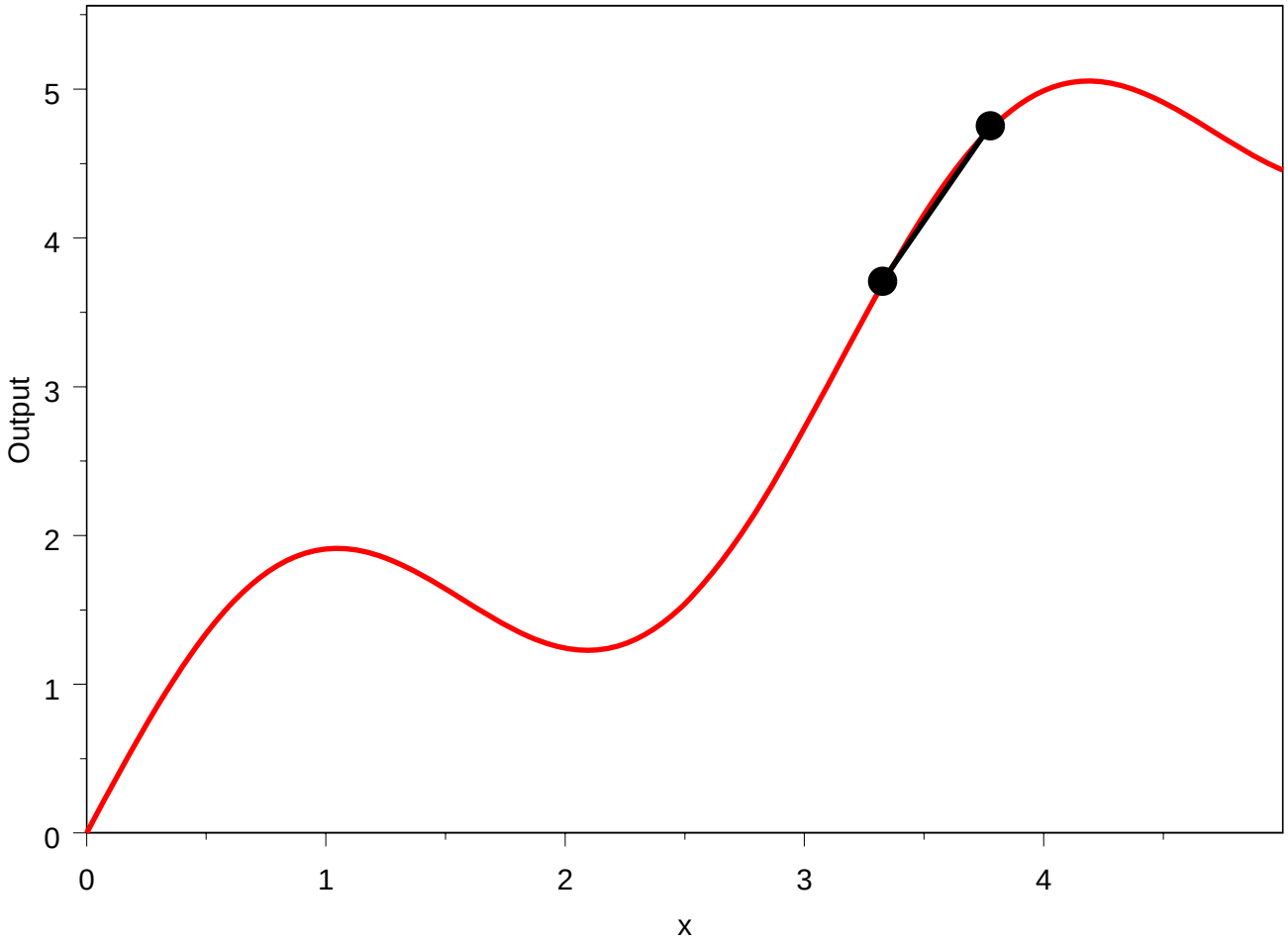
⋆ We have developed a screening method that is cheap and simple to implement.

⋆ It approximates gradient (elementary effects) in a one-at-a-time way to help us to identify nonlinear effects.

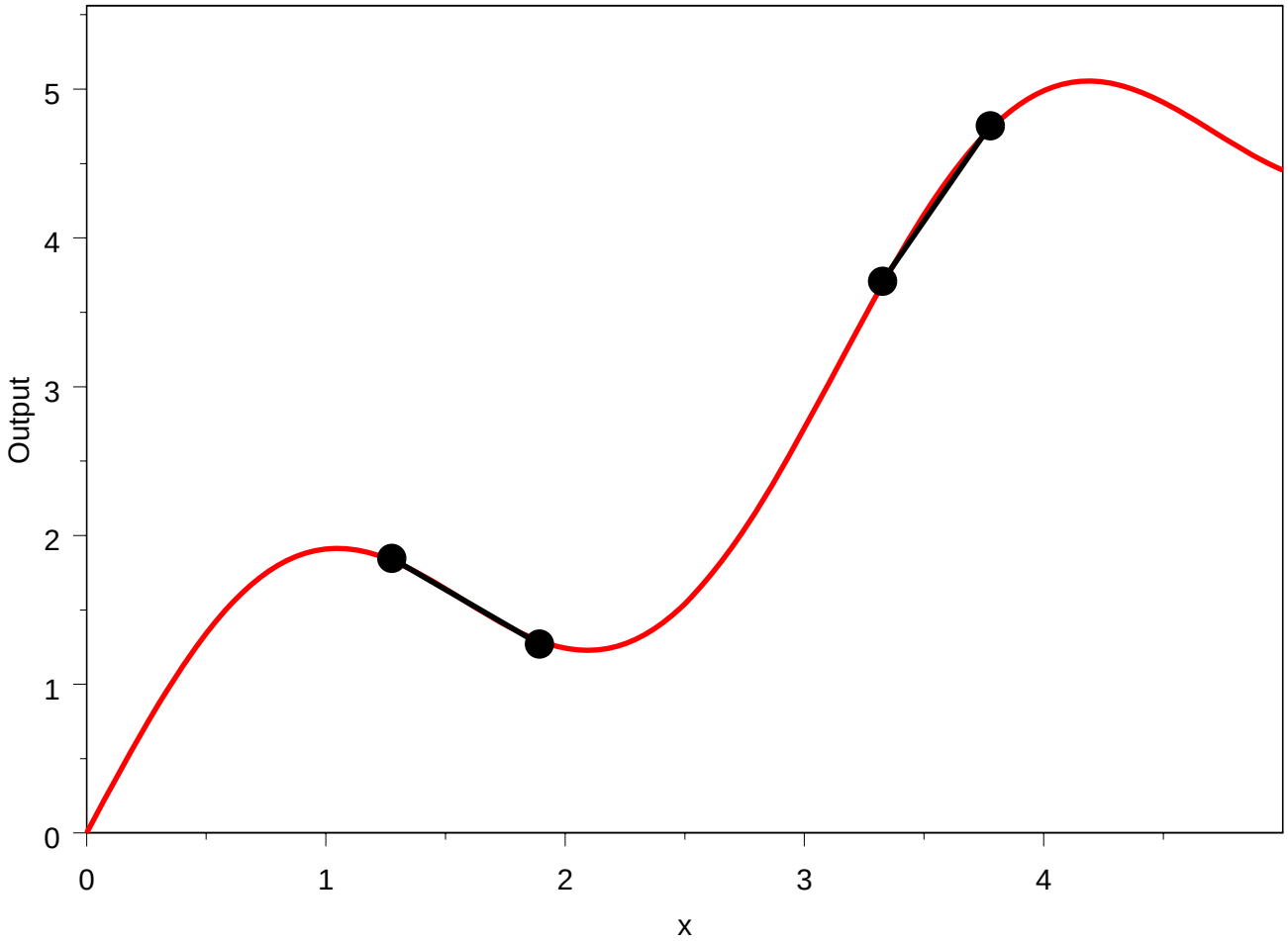
Screening



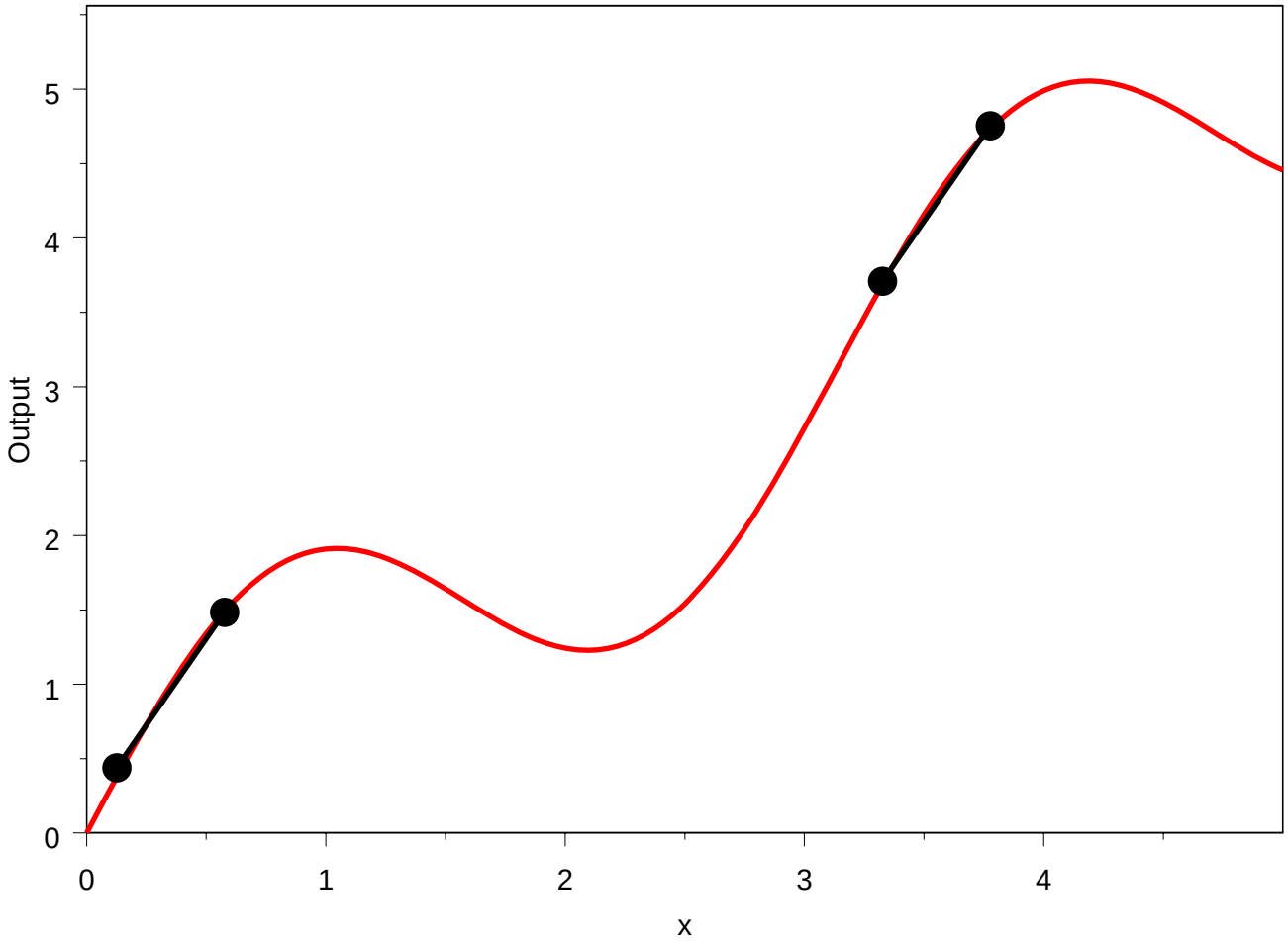
Screening



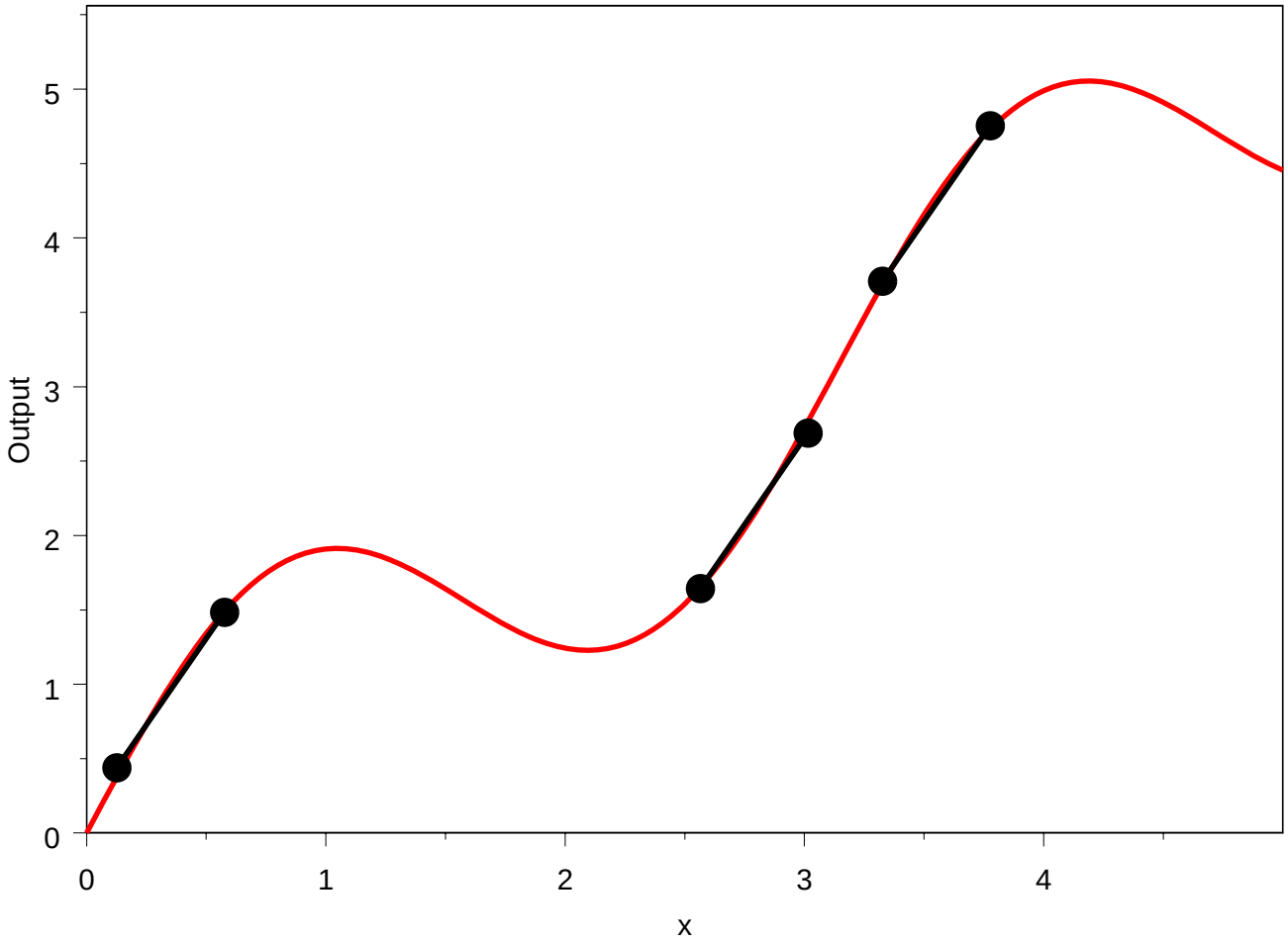
Screening



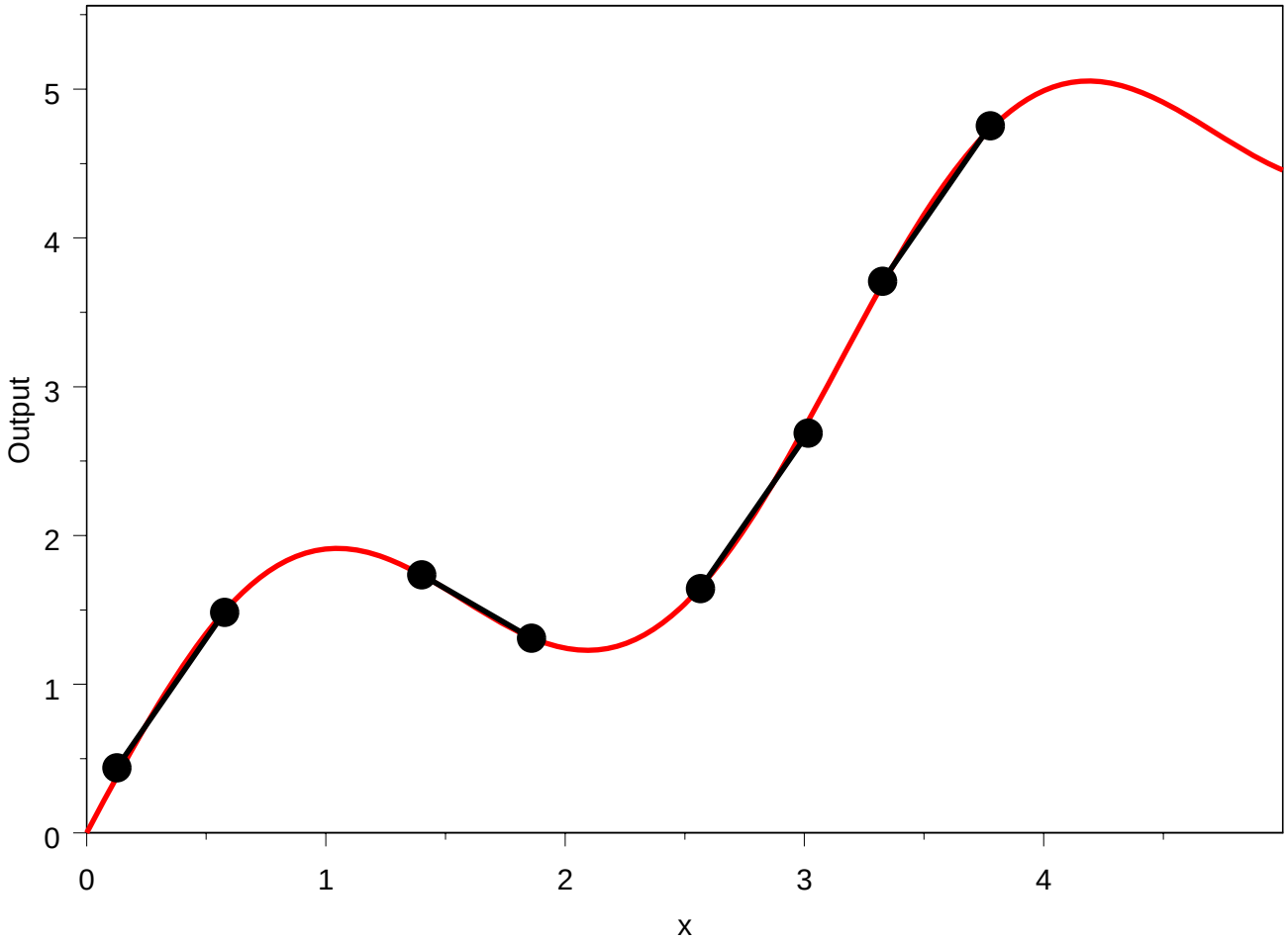
Screening



Screening



Screening



Sequential screening

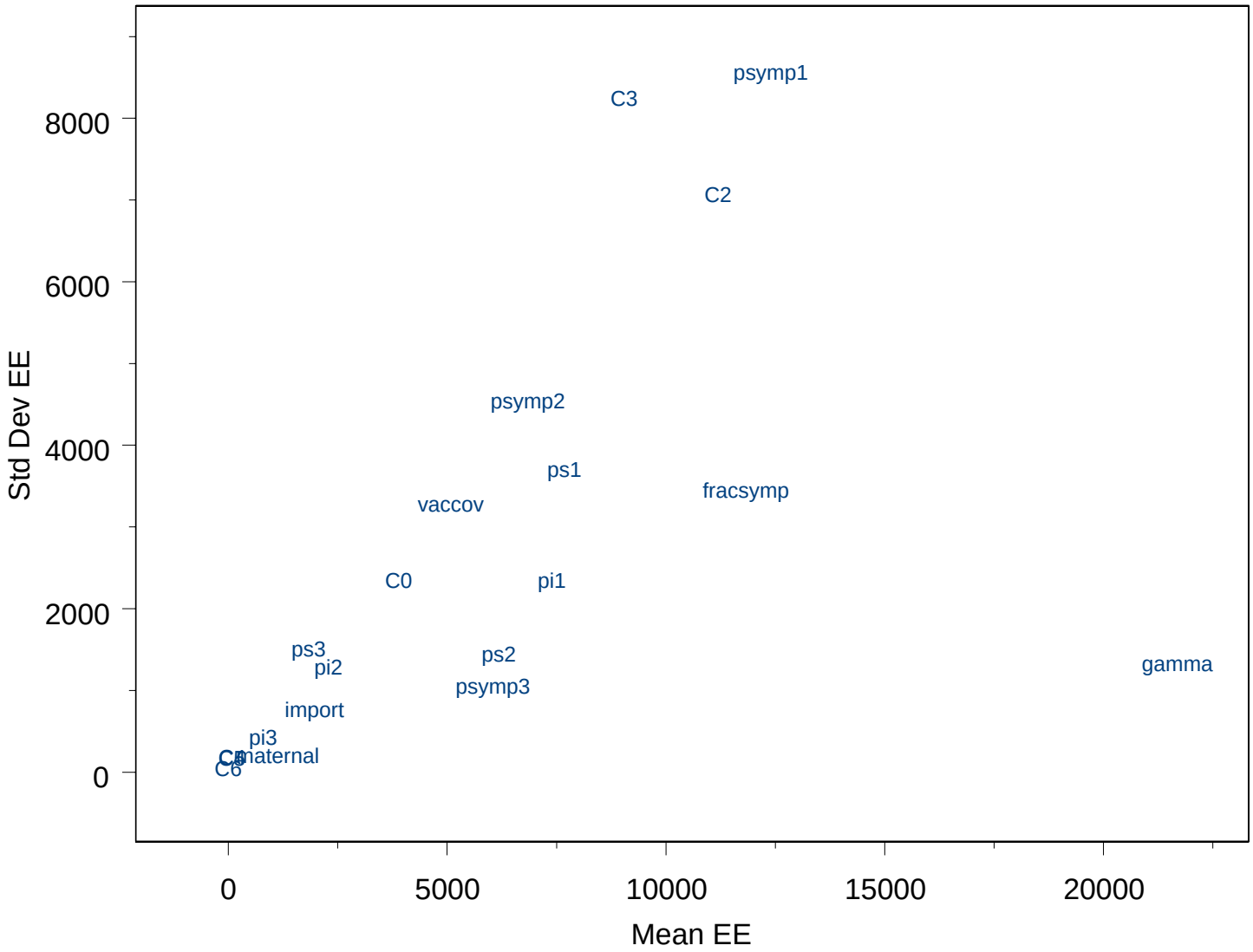
For each x_i, we start by fixing the rest of the inputs and get the model output for two different values of x_i.

We can use these two to calculate an elementary effect, EE_i, for x_i.

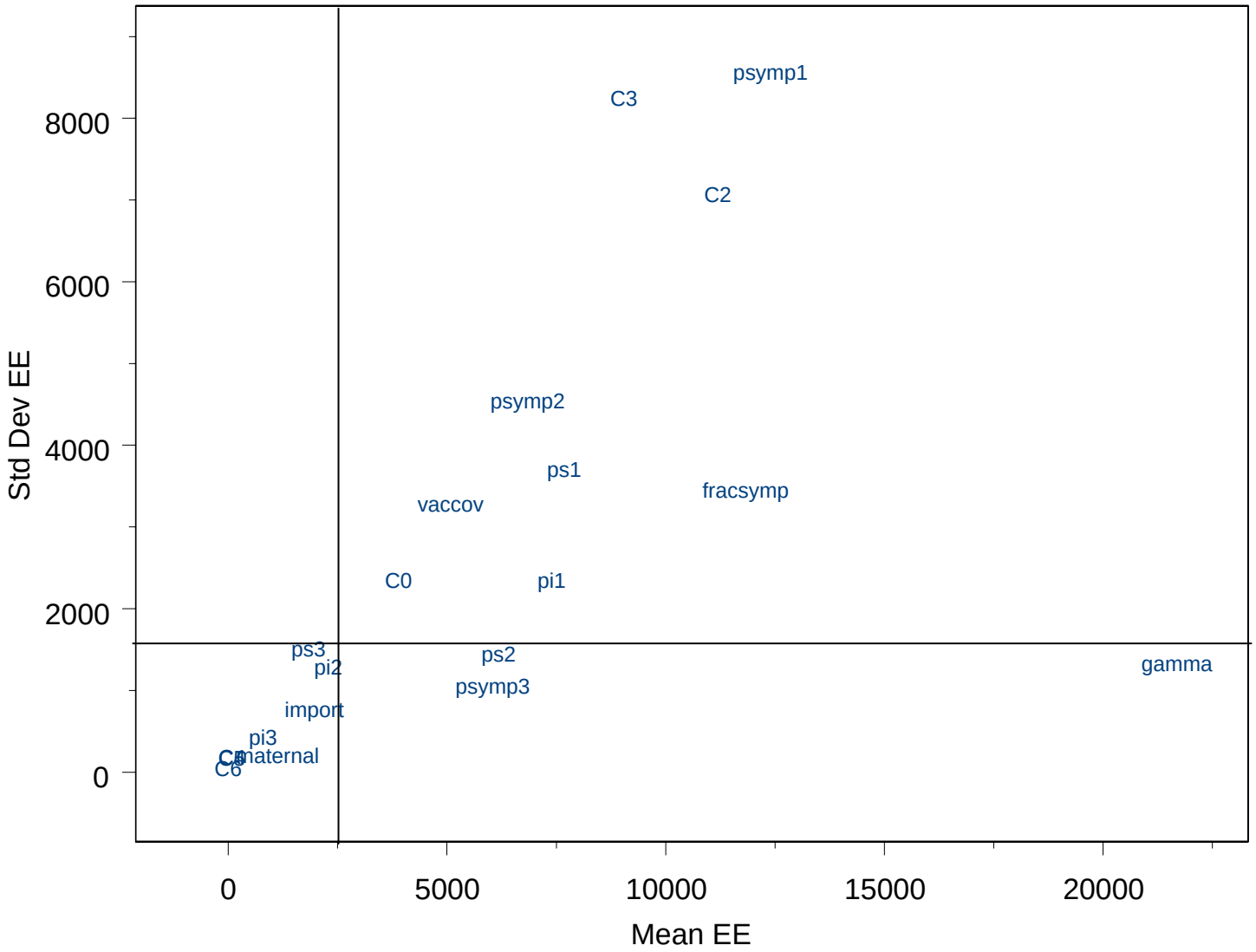
If, for x_i, we have done this twice, we can calculate the mean and variance of the EE_i.

If the variance is small, we continue our investigation and calculate more EE_i.

Results for the epidemiology model



Results for the epidemiology model



Probabilistic sensitivity analysis

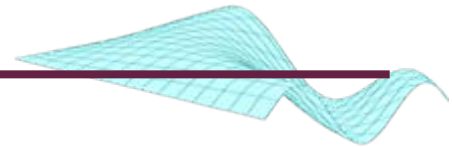
⋆ Global sensitivity analysis considers substantial changes in x_i.

⋆ How far to perturb individual x_i?

⋆ Perturbing each x_i to the limits could result in the range being unrealistically wide if many inputs are perturbed together.

⋆ These problems are overcome by acknowledging the uncertainty in x_i and treating it as a random variable with a specified distribution.

Elicitation



Elicitation is the process of turning someone's beliefs into a mathematical form for use in a statistical analysis.

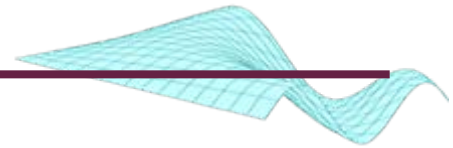
This process involves:

⋆ asking somebody questions about the quantity of interest,

⋆ mathematically modelling their beliefs,

⋆ checking that the model is consistant with their beliefs.

The bisection method



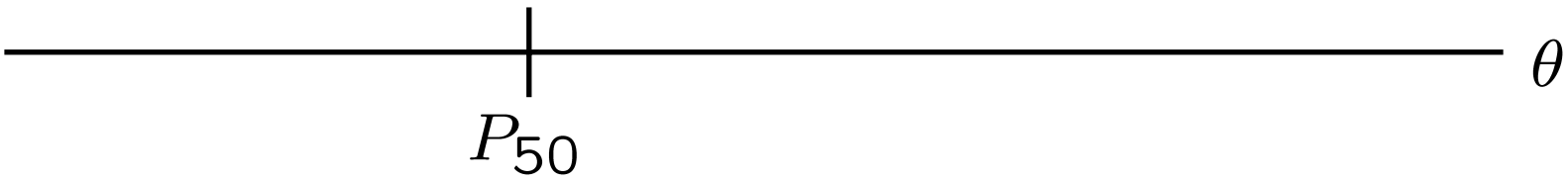
We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.

The bisection method

We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

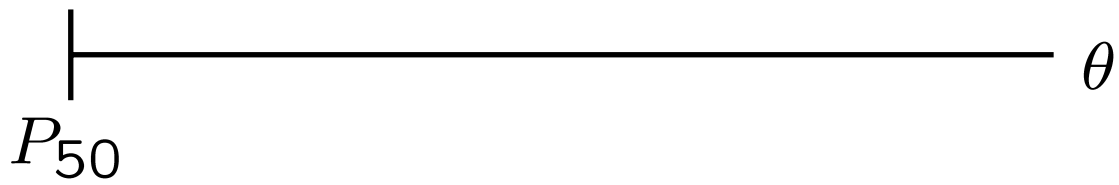
Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.



The bisection method

We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

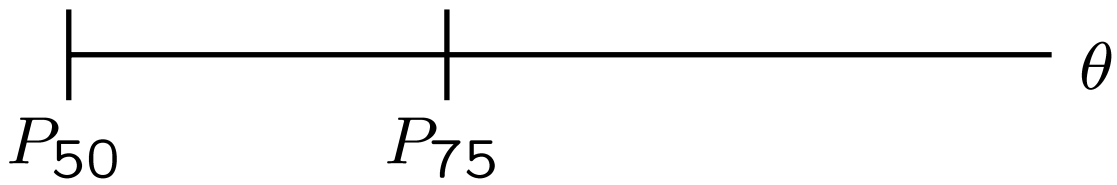
Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.



The bisection method

We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.



The bisection method

We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.

The bisection method

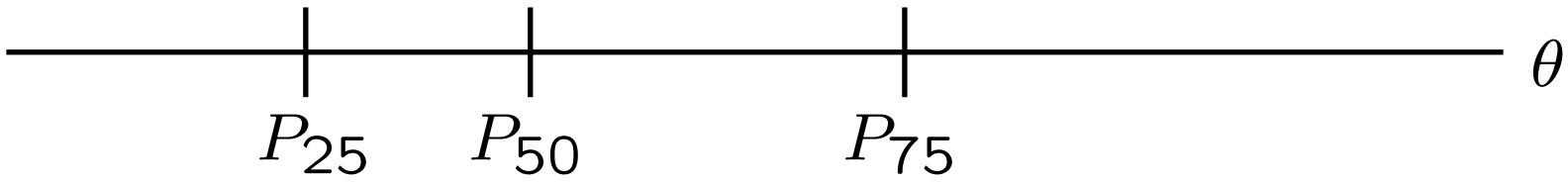
We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.

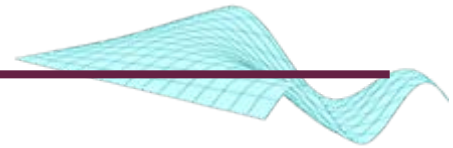
The bisection method

We want to find the expert's P_{25}, P_{50} and P_{75} for X_i.

Ask for a value that splits the real line into two intervals so that each has 50% chance of including the true X_i.



Elicitation



After a number of rounds of questioning and feedback, we are left with a probability distribution that characterises the experts uncertainty about the model inputs.

We use this distribution as the basis for our sensitivity analysis.

This enables us to focus on the important regions of the input space.

Main effects

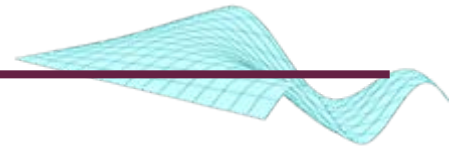
Some widely used methods of sensitivity analysis can be seen in terms of a decomposition of the function $\eta(.)$ into main effects and interactions:

$$y = \eta(x) = E(Y) + \sum_i z_i(x_i) + \sum_{i<j} z_{i,j}(x_i, x_j) + \ldots$$

where

$$\begin{aligned} z_i(x_i) &= E(Y|x_i) - E(Y), \\ z_{i,j}(x_i, x_j) &= E(Y|x_i, x_j) - z_i(x_i) - z_j(x_j) - E(Y). \end{aligned}$$

Variance decomposition

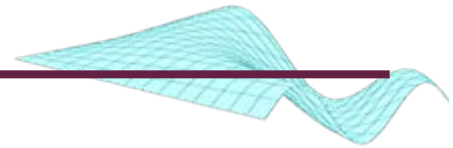


⋆ We can decompose the variance of the model output by considering the variance of the main and interaction effects.

⋆ This analysis of variance technique gives us a way of calculating how much each main effect contributes to the uncertainty in the output.

⋆ However, there is an assumption here that the inputs are independent.

Emulation



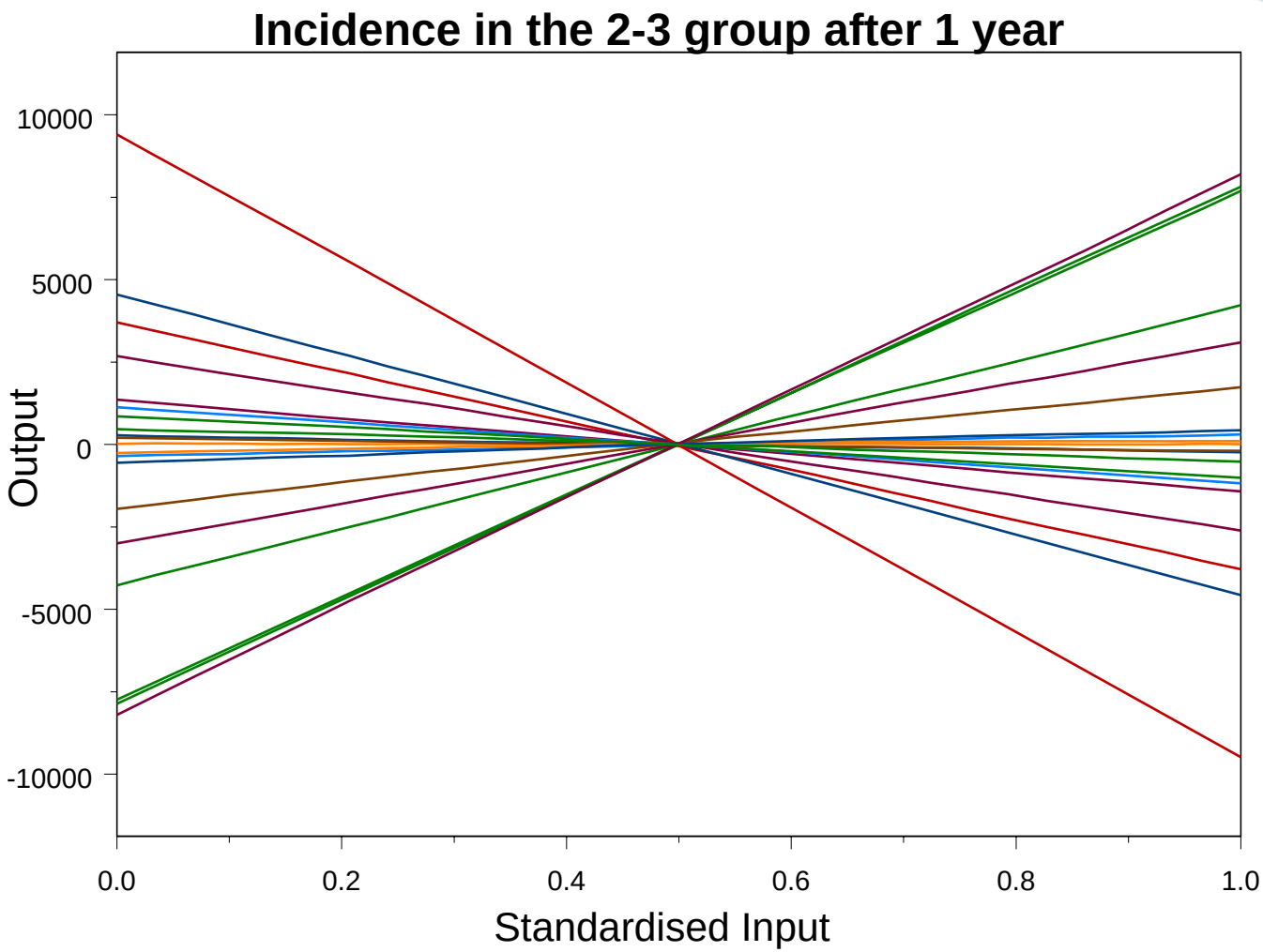
⋆ We could use Monte Carlo techniques to estimate these sensitivity measures.

⋆ This would cost us many thousands of model runs.

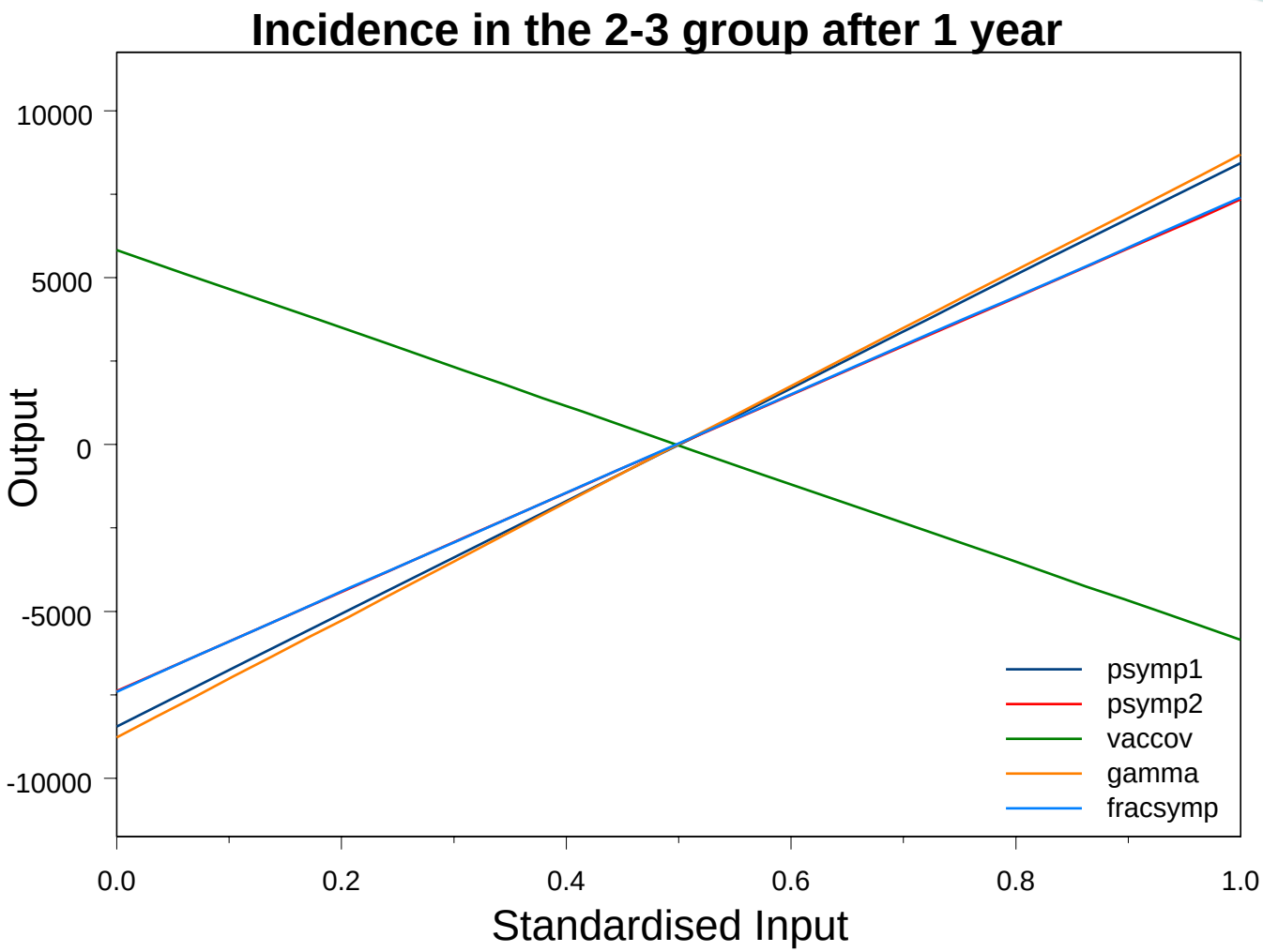
⋆ We can use the emulator technology to reduce this.

⋆ The subsequent results for the epidemiology model were based on just 242 model runs.

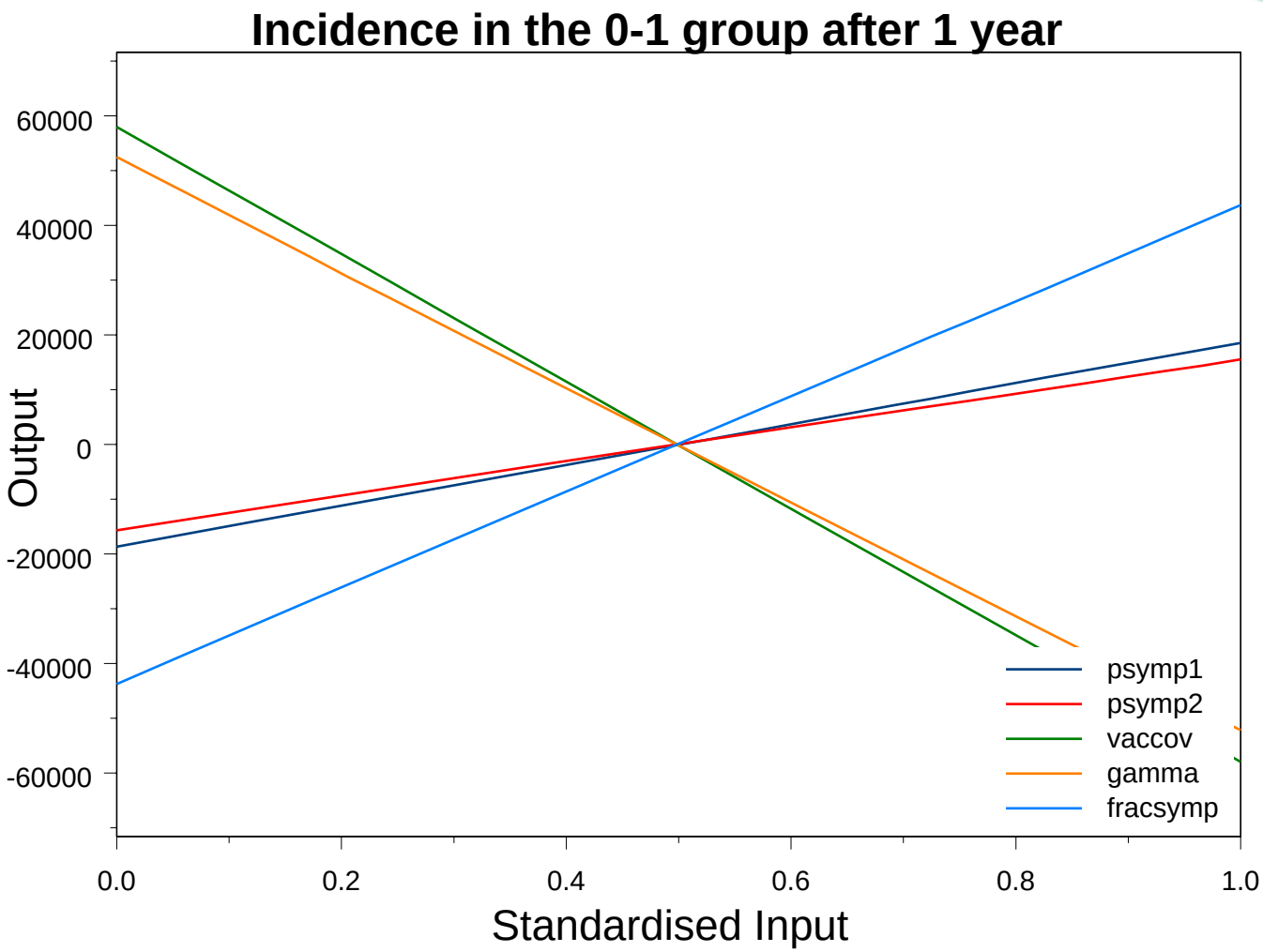
Sensitivity analysis results



Sensitivity analysis results



Sensitivity analysis results



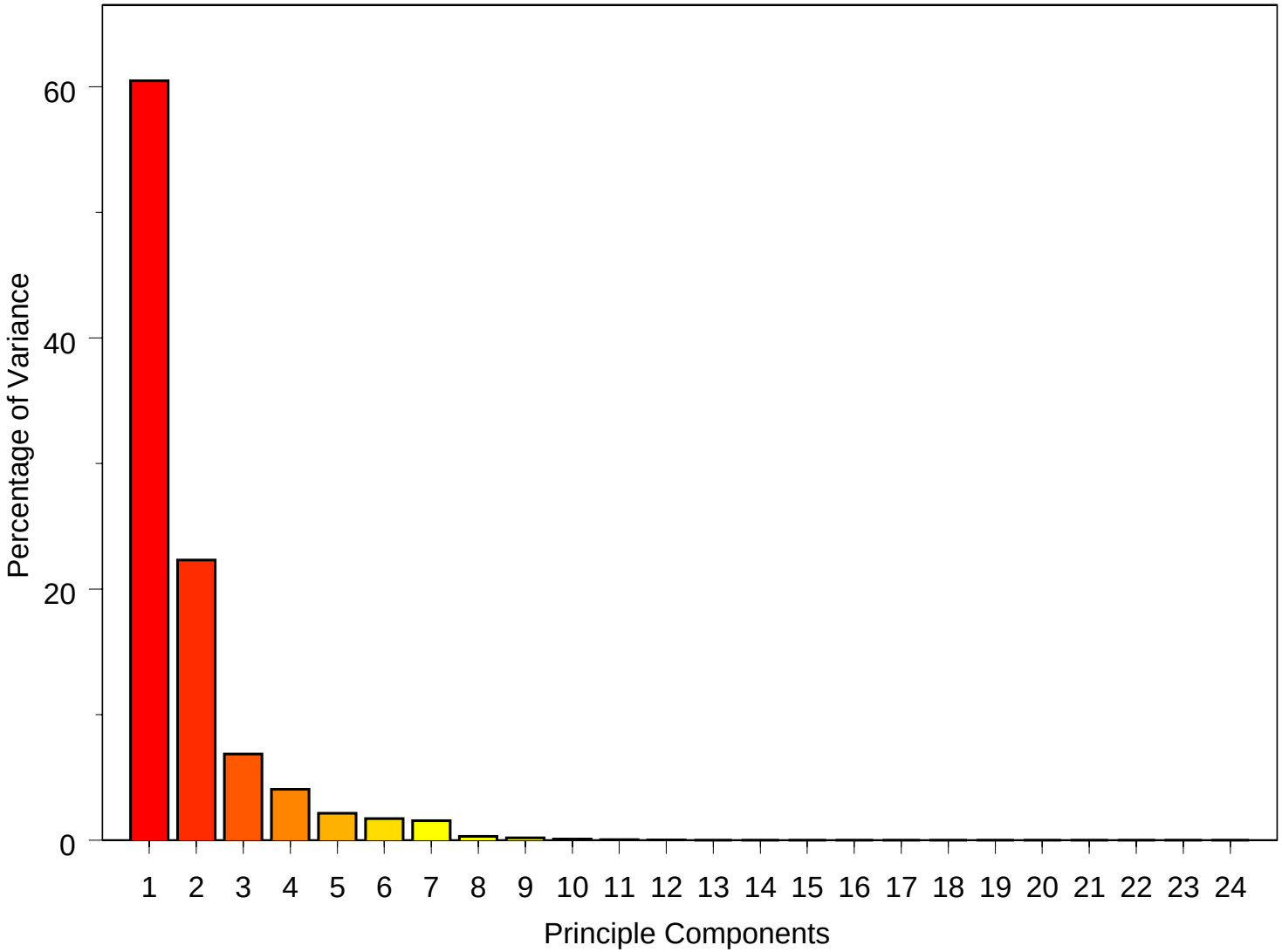
Handling multiple outputs

⋆ It is clear that the inputs are having varying effects of the incidence levels for the different age groups at different times.

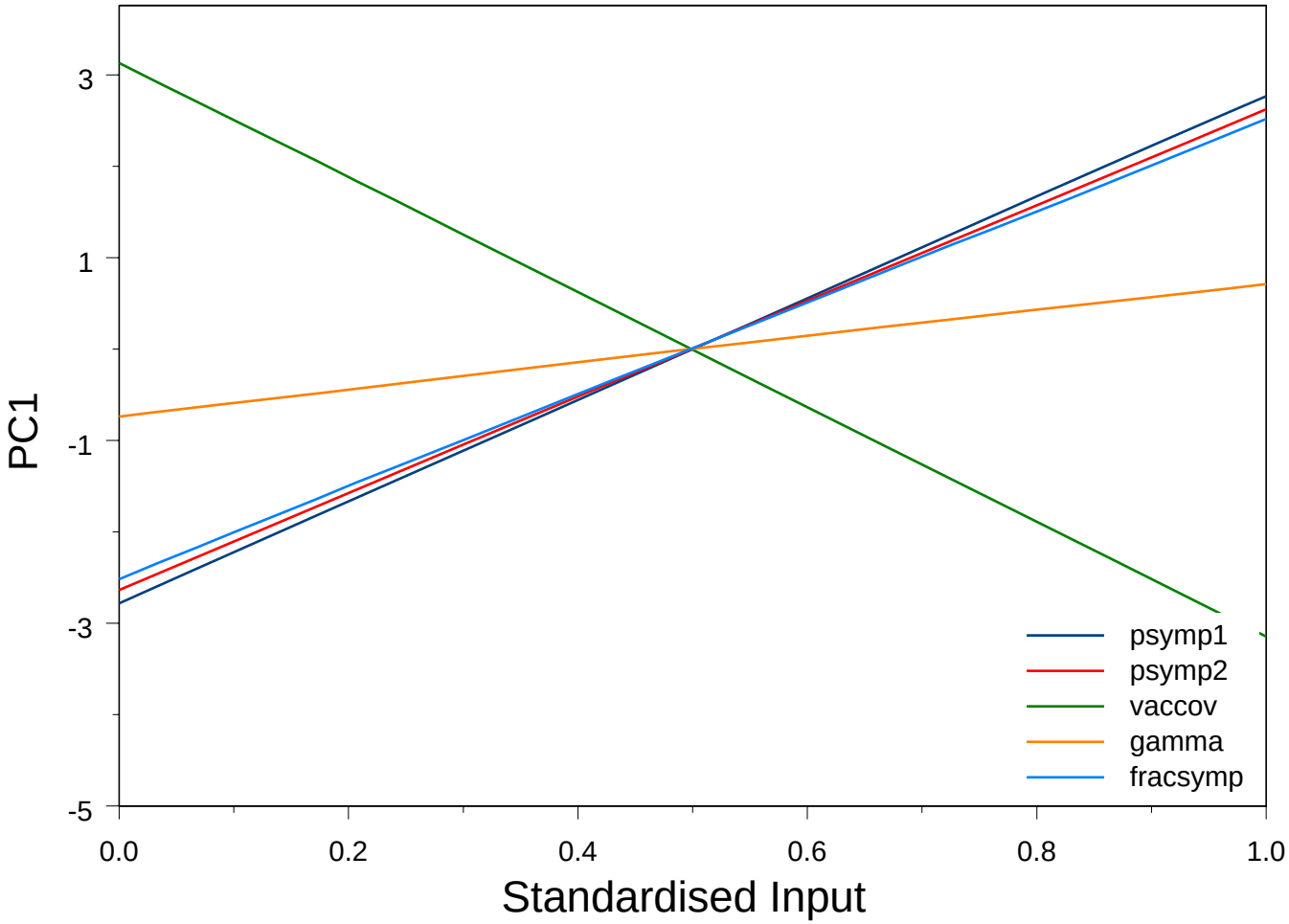
⋆ We could use a simple dimension reduction tool to gain an overview of the effect of the inputs on the incidence at all times of interest for all age groups.

⋆ Principle component analysis is simple to perform and maintains some interpretability.

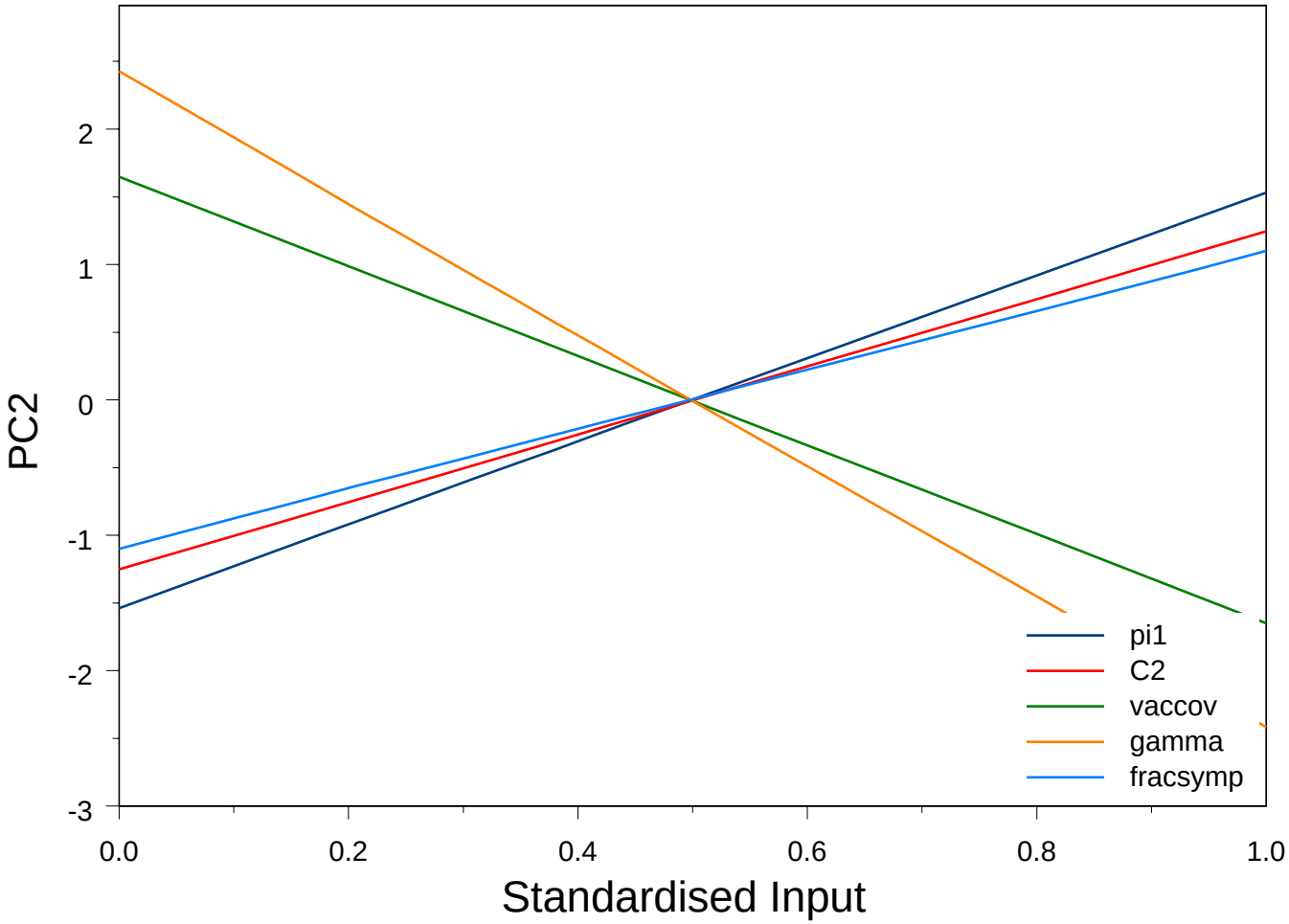
Handling multiple outputs



Handling multiple outputs



Handling multiple outputs



Variance contribution

	0-1 after 1	2-3 after 1	2-3 after 5	PC1	PC2
C0	1	0	0	0	2
C2	5	1	1	0	8
ps1	0	5.	3	6	3
ps2	0	4	4	6	2
psymp1	3	15	13	14	0
psymp2	2	11	12	13	0
pi1	12	2	0	0	13
vaccov	24	7	18	19	14
gamma	19	16	0	1	31
fracsymp	14	11	14	12	6
TOTAL	**89**	**83**	**75**	**82**	**86**

Figures correspond to the percentage of variance in the output explained by the input's main effect.

Variance contribution

	0-1 after 1	2-3 after 1	2-3 after 5	PC1	PC2
C0	1	0	0	0	2
C2	5	1	1	0	8
ps1	0	5.	3	6	3
ps2	0	4	4	6	2
psymp1	3	15	13	14	0
psymp2	2	11	12	13	0
pi1	12	2	0	0	13
vaccov	24	7	18	19	14
gamma	19	16	0	1	31
fracsymp	14	11	14	12	6
TOTAL	**89**	**83**	**75**	**82**	**86**

Figures correspond to the percentage of variance in the output explained by the input's main effect.

References

Gosling, J., Maruri-Aguilar, H. & Boukavalas, A. (2008). A conservative and efficient screening method for computer experiments. *To appear somewhere soon*.

Oakley, J. & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. R. Statist. Soc. Ser. B*, **66, 751-769.**

⋆⋆⋆ SHELF and GEM-SA ⋆⋆⋆