CLAHRC Wessex

Collaboration for Leadership in Applied Health Research and Care

NHS National Institute for Health Research

Arguments for and against the use of multiple comparison control in stochastic simulation studies

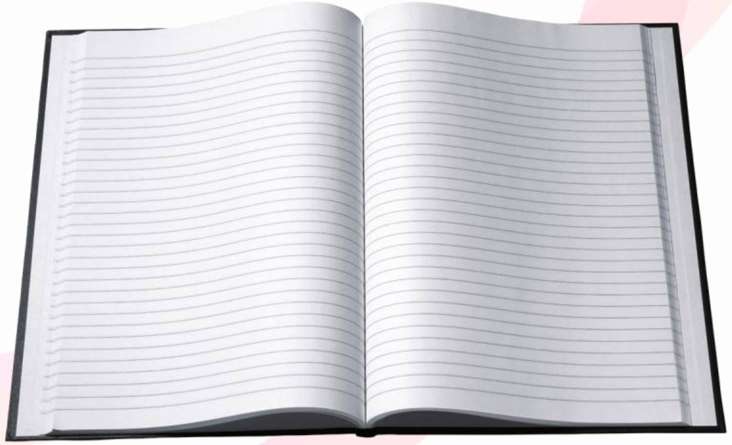
Thomas Monks. NIHR CLAHRC Wessex. University of Southampton.
Christine Currie. University of Southampton
Kathryn Hoad. University of Warwick

Simulation Workshop. April 2016.

UNIVERSITY OF Southampton

Motivation

- Pick up any standard Discrete Event Simulation (DES) text book and turn to the section on comparing scenarios



- They all talk about the Multiple Comparison Problem (MCP).
- But is a real issue in stochastic simulation studies?

What is the MCP? (1)

- It's about making mistakes in stochastic simulation studies!

- Mistakes in identifying differences between performance measures in competing configurations (scenarios) of a model when there are several or many alternatives to compare simultaneously

- Frequentist terminology: Type I or II error in inference.

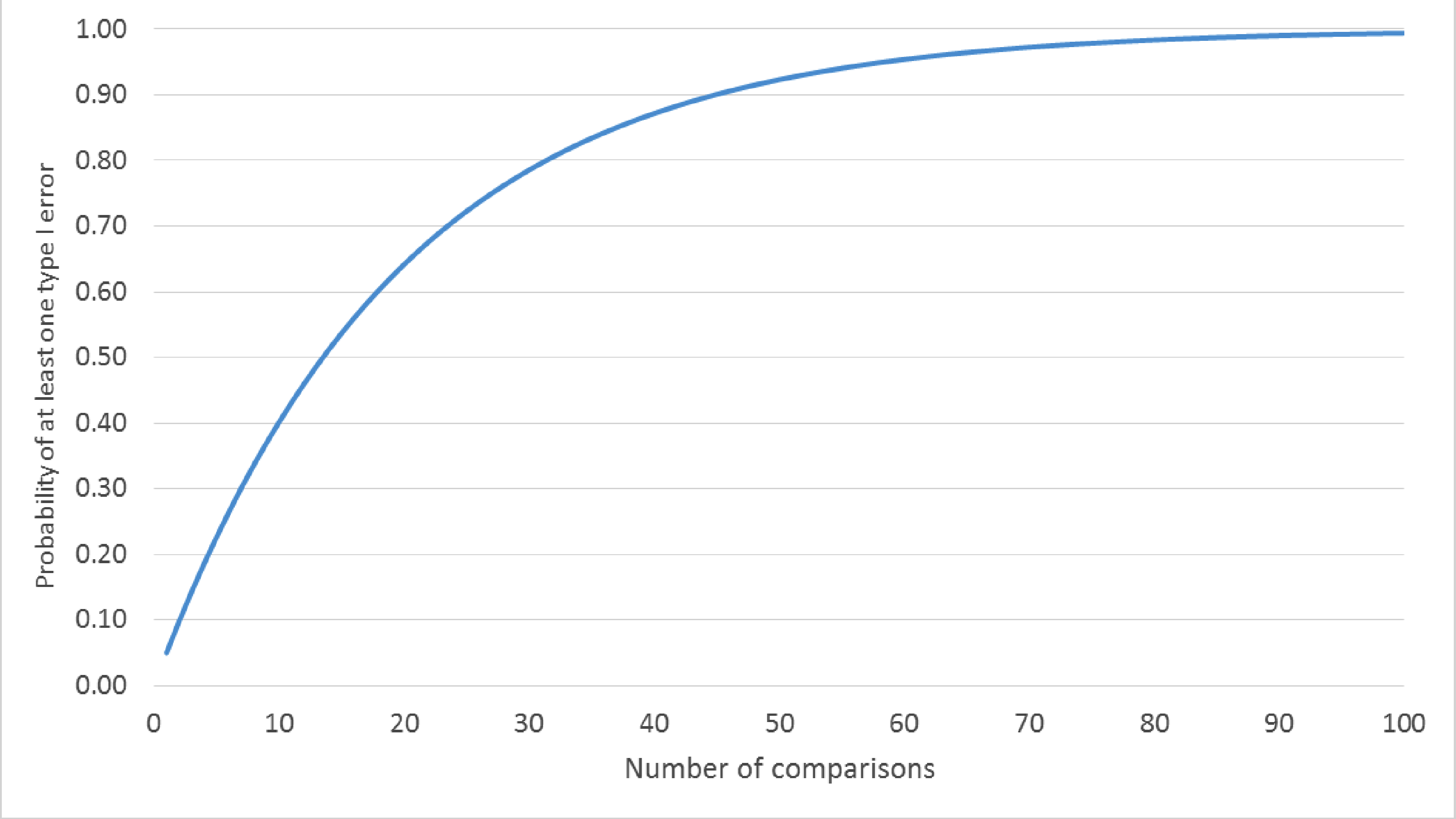
What is the MCP? (2)

- Two types of error rate
  - Per comparison error (e.g. 95% CI for scenario 1 versus scenario 2)
  - Overall experiment-wise error rate.

- We call the latter the **family-wise error rate (FWER)**
  - It is the probability of making a single Type I error in **m** comparisons

- When you conduct a large number of simultaneous comparisons the FWER inflates dramatically.

FWER Inflation



Alpha = 0.05.

Introducing Multiple Comparison Control (MCC)

- To tackle the MCP a number of MCC procedures have been developed.

- The most common MCC advocated in DES is the *Bonferroni Correction*.
  - This controls the FWER by creating stricter per comparison intervals.

- There are many alternatives, but in general the majority try to control FWER.
  - (we will introduce an alternative later).

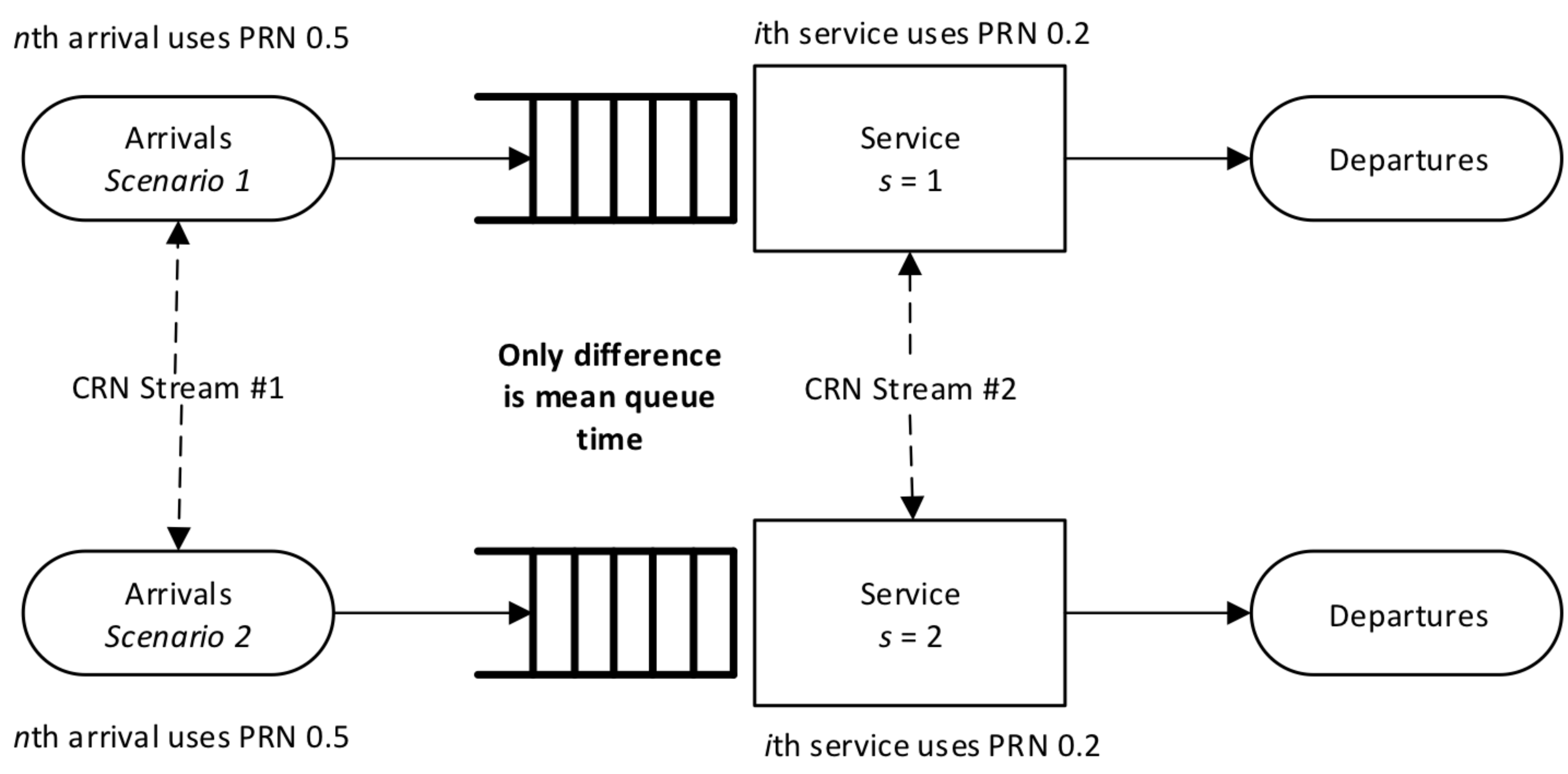
Arguments against the use of MCC

1. It often isn't practical

Scenarios	Comparisons (m)	α (Bonferroni)	CIs	
3	3	0.017	98.3	Interpretable, but wide CI's
4	6	0.008	99.2	
5	10	0.005	99.5	
6	15	0.003	99.7	Strict and CIs of little use
7	21	0.002	99.8	
8	28	0.002	99.8	
9	36	0.001	99.9	
10	45	0.001	99.9	

Bonferroni intervals = alpha / m e.g. 0.05/3 = 0.017 or a 98.3% CI.

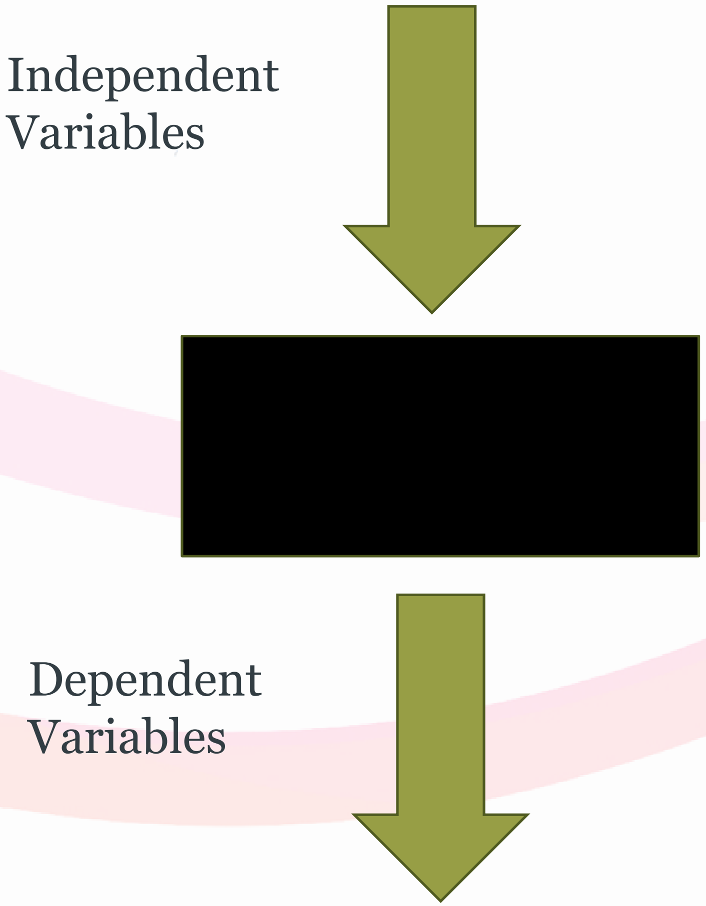
2. Common random numbers (CRN) "protect us"

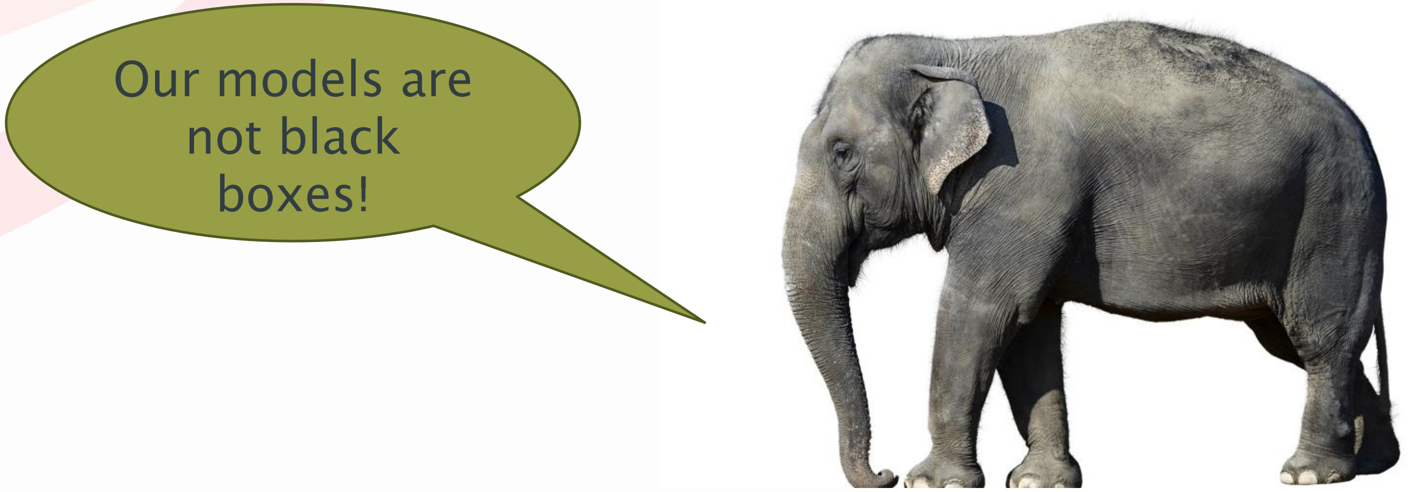


FWER cannot inflate at the rate predicted under these conditions.

3. Simulation is not a classical empirical study

- A big difference!
- We built the model.
- We can rerun the model.





Arguments for the use of MCC

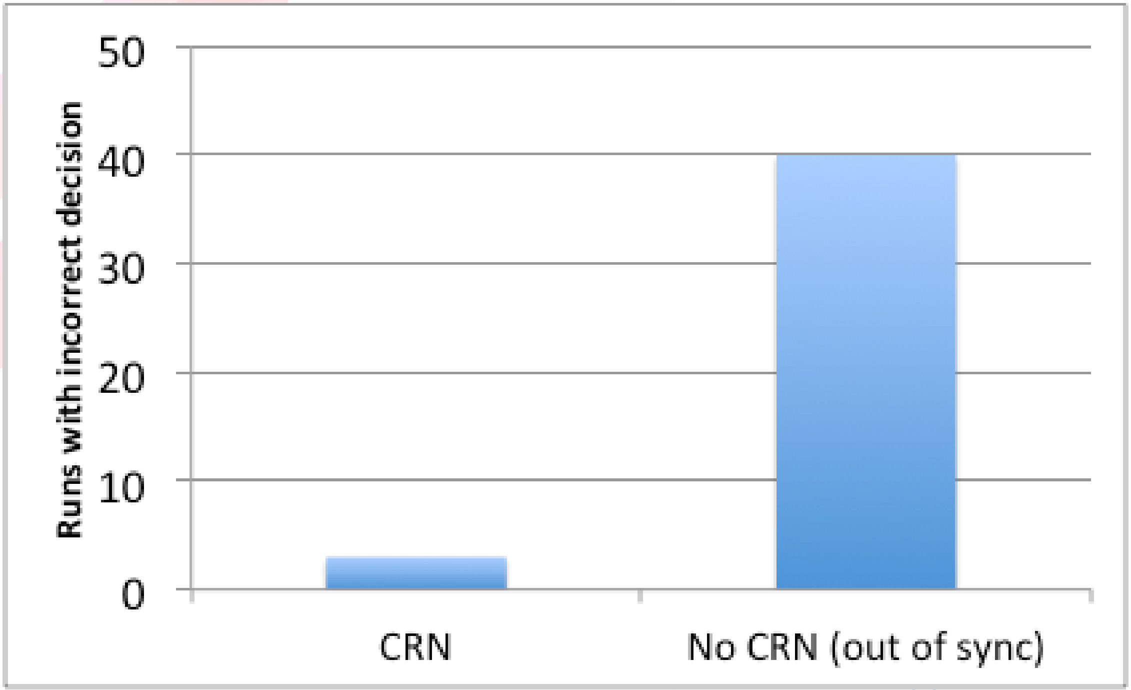
1. CRN do not always work!

- In complex models CRN might go out of sync. When this happens the MCP is a problem E.g.

- 100 replications of the following scenarios of an M/M/S system (adapted from Law (2006))
- Recorded the number of replications where the incorrect decision would be made.

Scenario	Parameters
M/M/1	$\rho = 0.9$ ($\lambda = 1$ min^{-1}; $\mu = 0.9$)
M/M/2	$\rho = 0.9$ ($\lambda = 1$ min^{-1}; $\mu = 1.8$)

- The correct decision is that the M/M/2 scenario has lower average queuing time.



2. An alternative to FWER control

- FWER control assumes that the role of the simulation study is to assist a decision maker is to pick the best system scenario out of n scenarios

- But complex models with many parameters are often used in an “exploratory” manner initially

- If the simulation study is more exploratory then control the **False Discovery Rate (FDR)**
  - Introduced by Benjamini and Hochberg (1995)
  - For algorithm and demo see Hoad and Monks (2011).

False Discovery Rates

Actual result	Decision		Total
	H_0 not rejected	H_0 rejected	
H_0 true	U	V	m_0
H_0 false	T	S	m_1
Total	m-r	R	m

V = the number of type I errors; T = the number of type II errors. Only m, r and m-r are observable. U, V, T, S and m_1 are unknown.

- Only m and R are known following a study
- FDR attempts to control E[V/R] (the expected number of Type I errors in all discoveries)
- As it is less “strict” you get less Type II errors (missing important results), but potentially some Type I errors (but less than if you have no MCC)

Final thoughts

- MCC is born out of the MCP in empirical studies.
- But does MCC really have a place in simulation studies as textbooks advocate?
- There are some compelling arguments that it is less of an issue, but mistakes can happen.
- For exploratory studies FDR procedures could be used over FWER control.
- If the goal of the study is to optimise then modellers should consider modern selection procedures (covered in paper) as opposed to MCC.

CLAHRC Wessex

Collaboration for Leadership in Applied Health Research and Care

NHS National Institute for Health Research

Thanks for listening

UNIVERSITY OF Southampton