Data Driven modelling of Antimicrobial Resistance (AMR) using Bayes Nets, regression and primary statistical analyses

David J Roser[a], Henrietta Venter[b], Sylvia Sapula[b], Michael Short[c]

[a]Water Research Centre, University of New South Wales, Australia

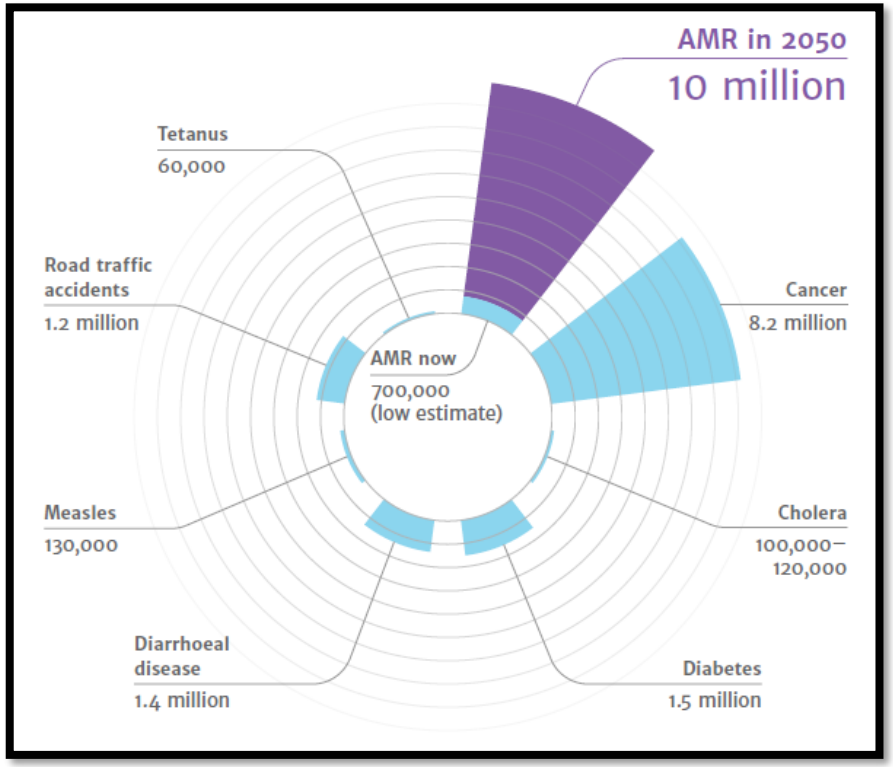
[b]Health and Biomedical Innovation, Clinical and Health Sciences, University of South Australia, Australia

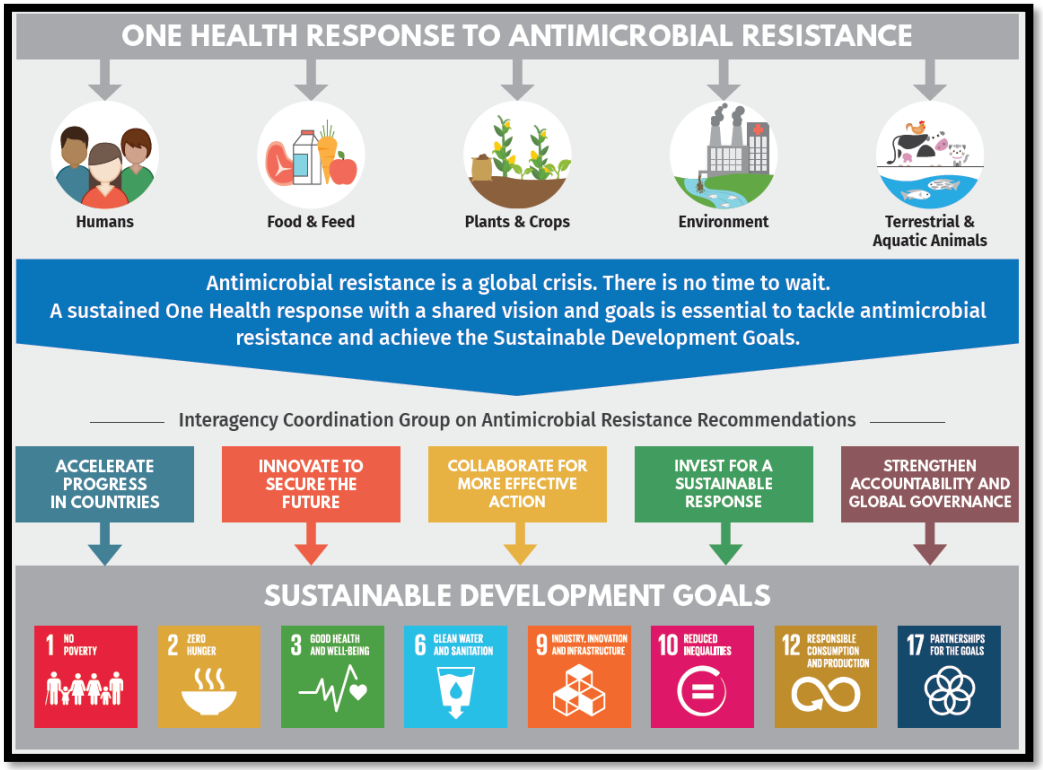
[c] University of South Australia, Future Industries Institute, Australia

[a]djroser@unsw.edu.au

The Antimicrobial Resistance Challenge

scale and complexity
despite 60 years of control efforts





O'NEILL, J. 2016. Review on Antimicrobial Resistance. Tackling drug-resistant infections globally; Welcome Trust and HM Government: London, England, .

IACG 2019. NO TIME TO WAIT: SECURING THE FUTURE FROM DRUG-RESISTANT INFECTIONS REPORT TO THE SECRETARY-GENERAL OF THE UNITED NATIONS. Interagency Coordination Group on Antimicrobial Resistance APRIL 2019. 28 pp.: World Health Organisation.

UNSW SYDNEY | Australia's Global University

University of South Australia

The most vulnerable population – the old?

Residential Aged Care (RACF) AMR Project needs: quantitative systems & risk analysis

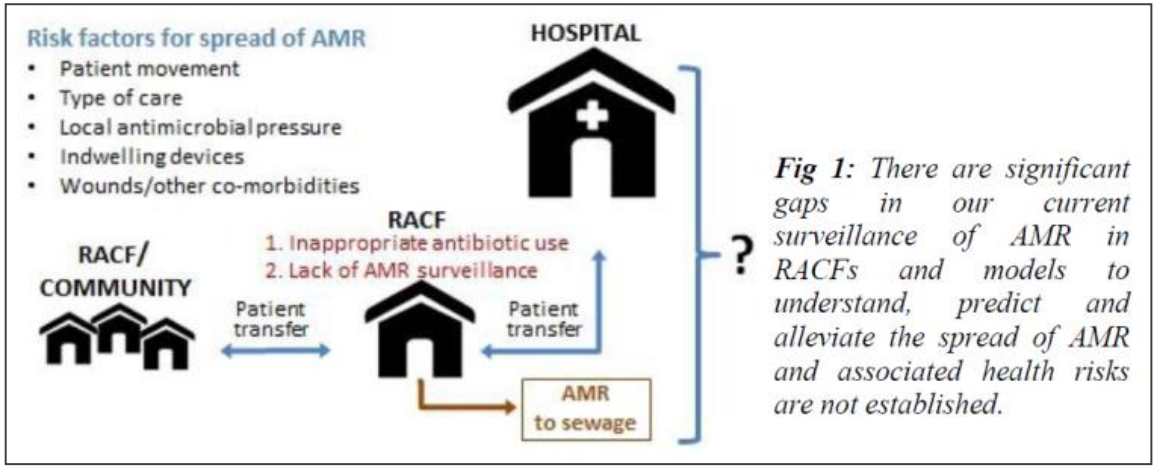
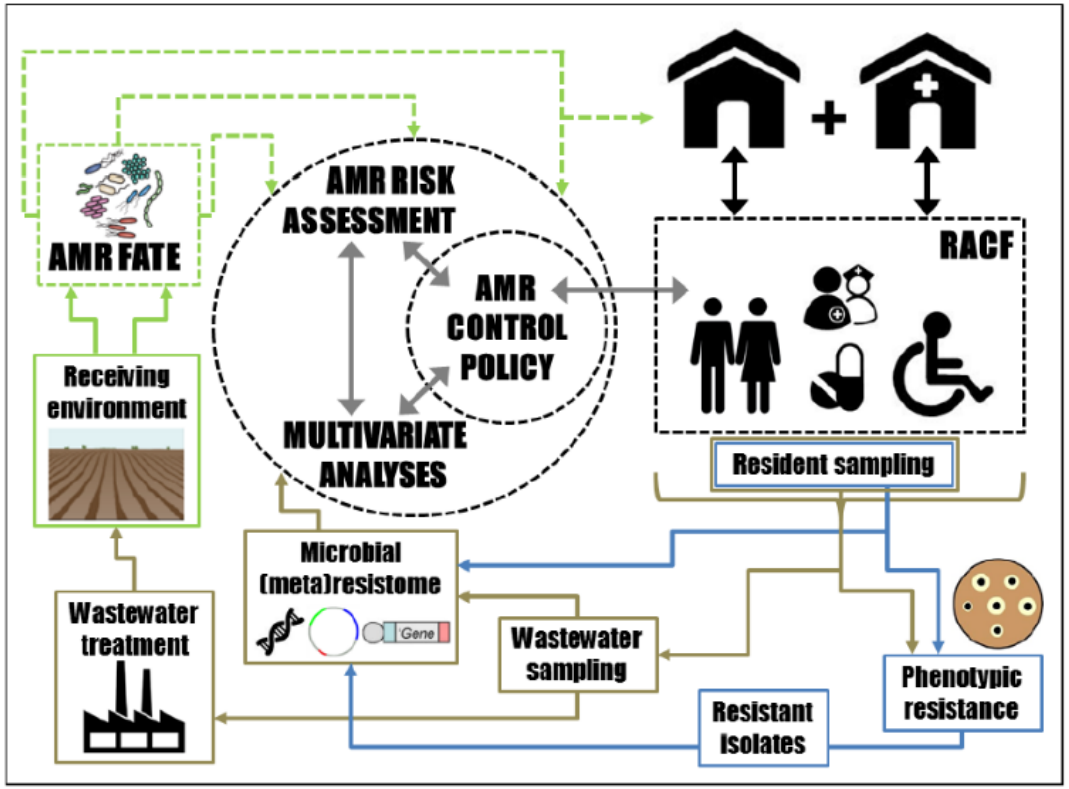


Fig 1: There are significant gaps in our current surveillance of AMR in RACFs and models to understand, predict and alleviate the spread of AMR and associated health risks are not established.



Aims-How?

RACFs most vulnerable subgroup?

1. **Adapt quantitative microbial risk assessment (QMRA = HHRA)**

2. **Bayesian Inference > "eco-molecular" framework**

3. **Bayesian Belief nets > operational framework (review, data mining, experimental program, integration, risk management)**

"Conclusions: We propose that environmental aspects of antibiotic-resistance development be included in the processes of any HHRA addressing ARB. **Because of limited available data, a multi-criteria decision analysis approach would be a useful way to undertake an HHRA of environmental antibiotic resistance that informs risk managers**."

ELLISON, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications, 6, 1036-1046.

CITATION: Ashbolt NJ, Amézquita A, Backhaus T, Borriello P, Brandt KK, Collignon P, Coors A, Finley R, Gaze WH, Heberer T, Lawrence JR, Larsson DG, McEwen SA, Ryan JJ, Schönfeld J, Silley P, Snape JR, Van den Eede C, Topp E. 2013. Human health risk assessment (HHRA) for environmental development and transfer of antibiotic resistance. Environ Health Perspect 121:993–1001; http://dx.doi.org/10.1289/ehp.1206316

Methods 1 – Variables/nodes

Survey Data from RACFs (Genomics??)

54 variables (True/False, numbers, ratings, durations)

- 4 subpopulations, 3 RACF and 1 independent residential
- Age, health status, gender, hospitalization
- True/False & isolates/person
  - E. coli (single, multidrug, total)
  - MRSA
  - VRE
  - Total AMR isolate incidence
- RTI/UTI
- Emergency/Hospital stays
- Antimicrobial course details number, duration, total time
- Systemic v. Topical application

Antimicrobials (T/F)

- Fluoroquinalones, Betalactam, Cephalosporin, Eardrops, Amoxycillin, Augmentin, Cefalexin, Chloramphenicol, Ciproxin, Doxycycline, Nitrofurantoin, Trimethoprim

Methods 2 – Data analysis and modelling

Data driven modelling

- Screening of AMR drivers: Descriptive stats, Correlation
- 'Model' construction: Regression, **Semi-naïve Bayes nets (Netica v 6.07)**
- Process of elimination of low contribution drivers of AMR
  - Select variable with higher R compared to AMR related variables;
  - Residue explored and further trimmed: Use backwards probability based regression; **Eliminate factors with variance reduction or mutual information < 1% in sensitivity to findings**
  - **Identify provisional drivers and data robustness**.

SIRS simulation – 3-10 years simulations for comparison.

Summary statistics

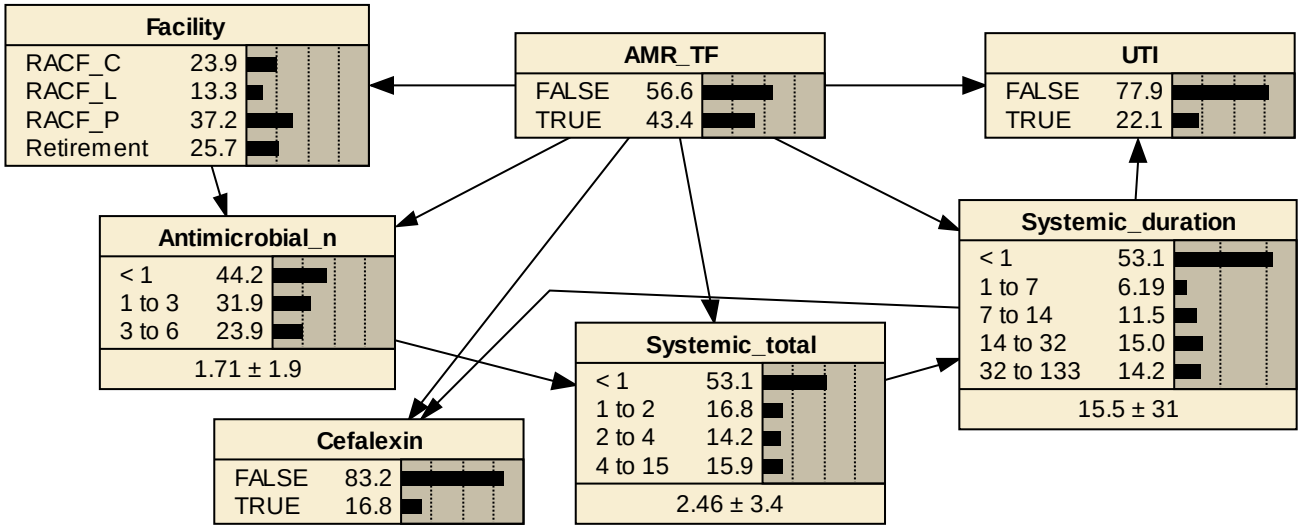
- 84 RACF residents (mean 87 yo), 29 independent living in community (mean 78 yo)
- 56% received antibiotic courses in 2020
- **% AMRs Detected = 43% of residents studied, 39% E. coli**
- **45% of RACF residents' carriers, 38% of independent living were carriers**
- **21% RACF carried Multi Drug Resistant E. coli v. 14% of community group**
- Only single Vancomycin resistant Enterococcus isolate
- Most Correlation R values < 0.2 especially Spearman rank matrix = no dominant driver

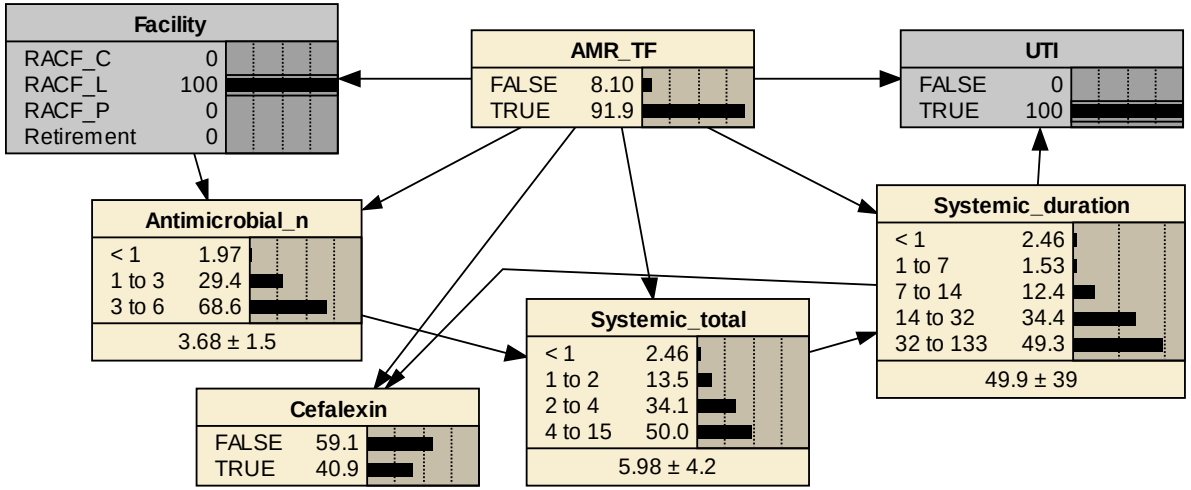
Most significant candidate variables

Total AMR/ E. coli variables	MRSA
RACF (which)	RACF (which)
UTI (T/F)	UTI (T/F)
RTI (T/F)	RTI (T/F)
Systemic_T/F	Systemic_T/F
Antimicrobial_n	Antimicrobial_n
Systemic_duration(days)	Systemic_duration(days)
Systemic_total_n	Systemic_total_n
Cefalexin T/F	Cefalexin T/F
Dementia_rating (0-3)	Female_T/F
Antibiotics_T/F	Mobility_rating (0-3)
Hosp_stay (0-2)	BenzoD_T/F
	betalactam T/F
	Doxycycline T/F

Bayes Net exploration

Total AMR (T/F) example UTI in Facility L





Group presentation & exploration of variation e.g. between facilities

Also interrelationships between antimicrobial dosing evident in RACFs

Bayes v. Regression 1 – Total antimicrobial resistant isolate incidence (n)

Bayes Net (% Variance reduction)		Backwards Regression (coefficient P.)	
Systemic_duration	**9.17**	**Systemic_duration**	**0.0013**
Systemic_total	**6.72**	**Systemic_total**	**0.0968**
Facility	6.61	**Antimicrobial_n**	**0.1576**
UTI	**4.37**	**UTI**	0.14
Antimicrobial_n	**4.18**		
Cefalexin	3.7	(overall P. <0.0001 and R^2 = 0.29)	

('Facility' variable not suited to regression but fine for BNs)

Bayes net v. Regression 2.
Multidrug Resistant E. coli incidence (n)

Bayes Net (% Variance reduction)		Backwards Regression (coefficient P.)	
Systemic_duration	**9.38**	**Systemic_duration**	**<0.0001**
Systemic_total	6.86	RTI	0.0702
Facility	5.66		
Antimicrobial_n	2.86		
UTI	2.17	(overall P. <0.0001 and R^2 = 0.22)	
Age2020	1.44		
Hosp_stay	1.16		

('Facility' variable not suited to regression but fine for BNs)

Variables having low to trivial influence

Gender (women more vulnerable to UTIs?)

Age (age vulnerability?)

Hospital admission (transfer from antibiotic rich hospital environment?)

Statins (circulation problems?)

Influenza vaccination (vaccine sensitivity?)

Topical applications (route of exposure of AMR - v. systemic)

Beta lactam antibiotic use (Amoxycillin, Augmentin)

Most major antibiotics including fluoroquinalones, cephalosporins, chloramphenicol, doxycycline

Major summary findings

(Within the limits of sample sizes and detection frequency)

- **Regression and BNs drivers similar there statistical significance of causal coefficients is high/marked (P.<0.01)**
- **The biggest issue was *E. coli* AMRs (implication for genomics studies)**
- **Relationships statistically significant but still not strong overall : 0.2-0.3 r^2**
- Antibiotic T/F v. AMR bacteria Correlations = 0.196/0.132 (Spearman/Pearson)
- Main overall driver appears to be high systemic doses of antibiotics (duration/number of courses, total treatment time = selection pressure)
- Marked difference between RACF incidence (independent living intermediate)
- E. coli single resistance not strongly correlated to any drivers but multi drug resistance corresponds to high dosing
- VRE was not a major issue (past major concern) however single detect was with most extensive dosed resident (161 days and antibiotic treatment)

UNSW SYDNEY | Australia's Global University

University of South Australia

What is missing?

A major node/variable? (unknown)

Or

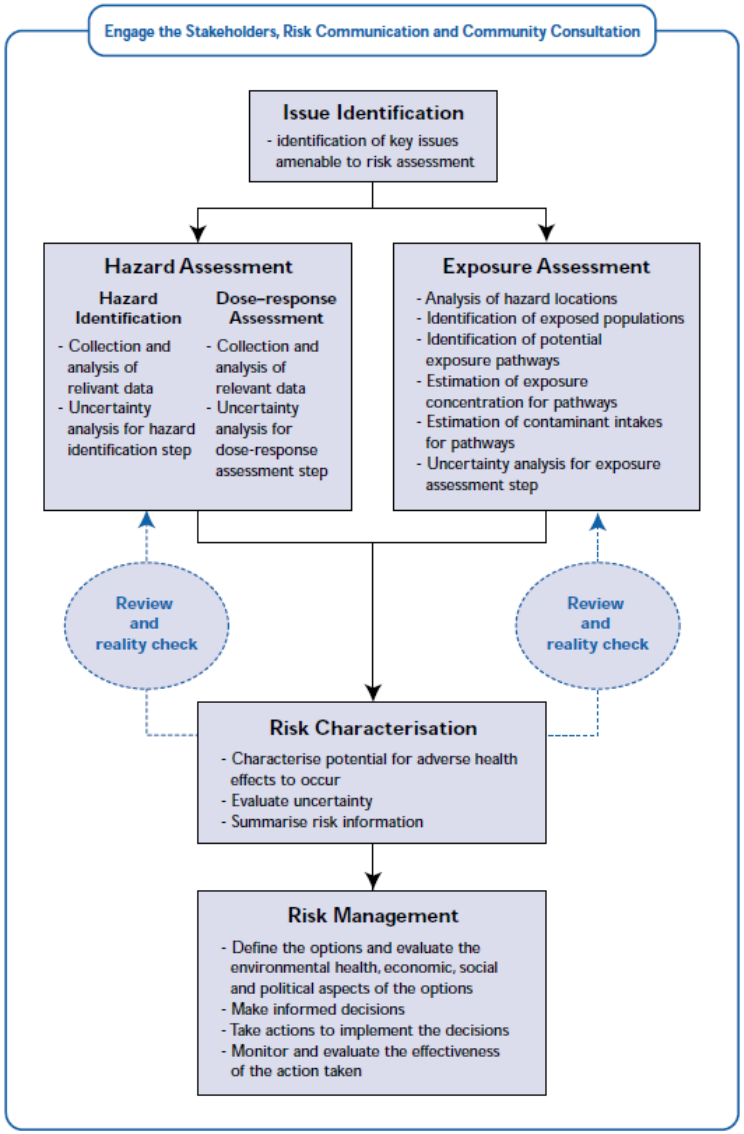
Insufficiently sensitive testing? (method)

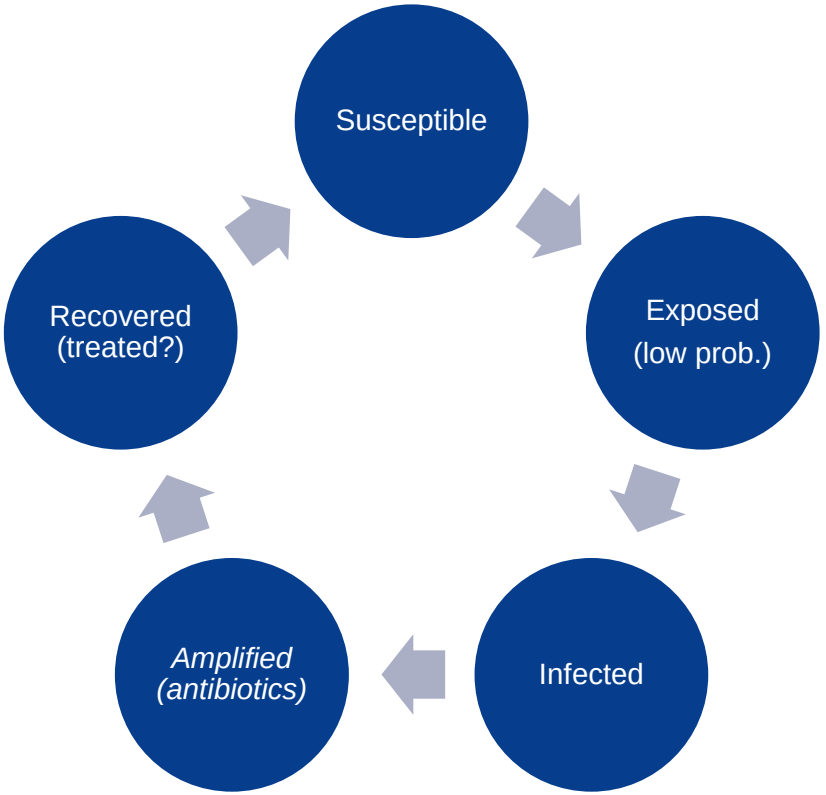
Or

Is something else going on?

QMRA v. (causal/linear)

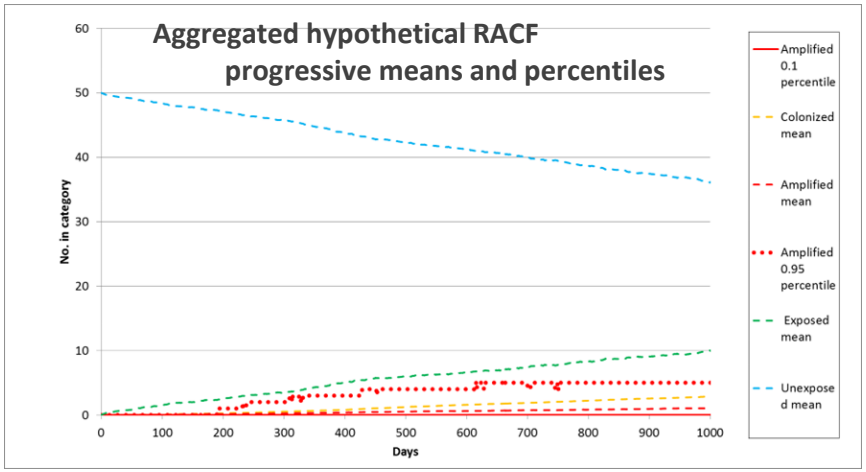
Epidemiology (SEIR/cyclic modelling)

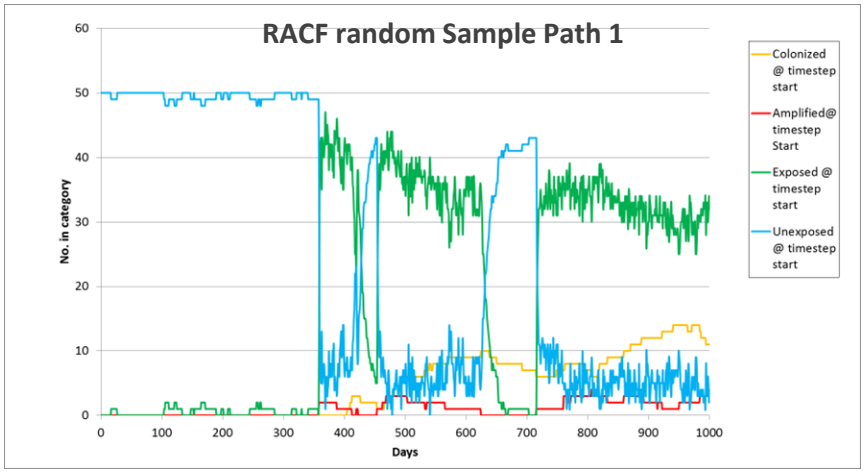


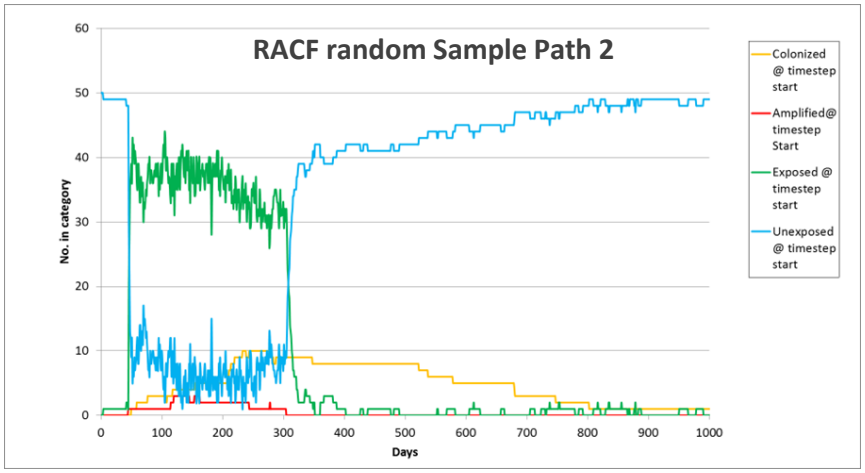


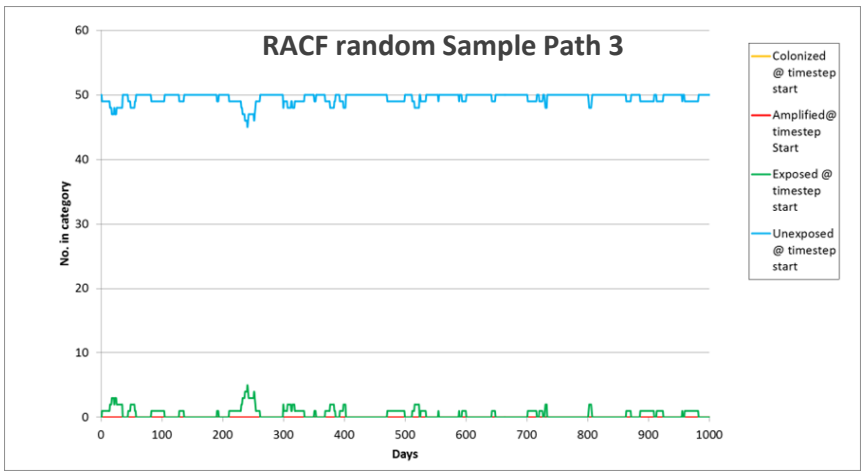
Infrequent but high impact events a challenge for dynamic BNs

The impact of SEIR stochasticity





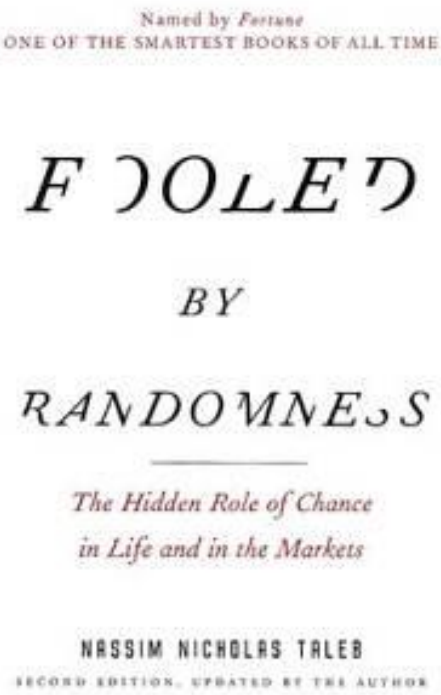




n=50, 3 years, ρ (prescription induced amplification rate = 0.0008/person /day)

Parting thought – Are causal BNs fooled by randomness?

- **Need to integrate Data Driven & Process Based modelling**
- **Value of bringing more data?**
- **KPI comparison of facilities (Business management SOP)**
- Is unaccounted for variance Bayesian Net due to:
  - Unknown drivers?
  - Noisy primary input data?
  - **AMR outbreak evolution induced feedback cycles? (we expect varying RACF to RACF status quos)**
- Colonized v. Amplified (level of 'infection')
- Are statistically significant relationships actually more significant than metrics suggest?



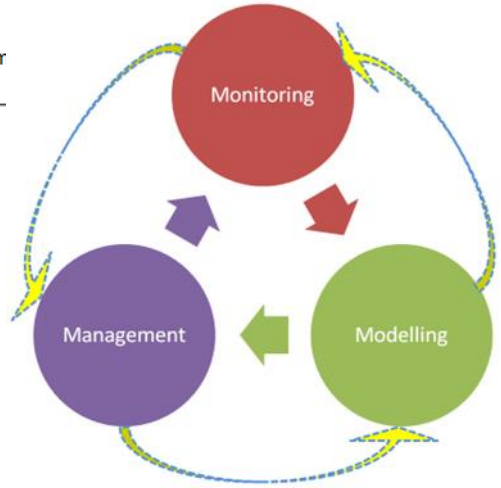
Acknowledgements

NHMRC funding to Rietie Venter Uni SA

The monitoring, modelling and management cycle

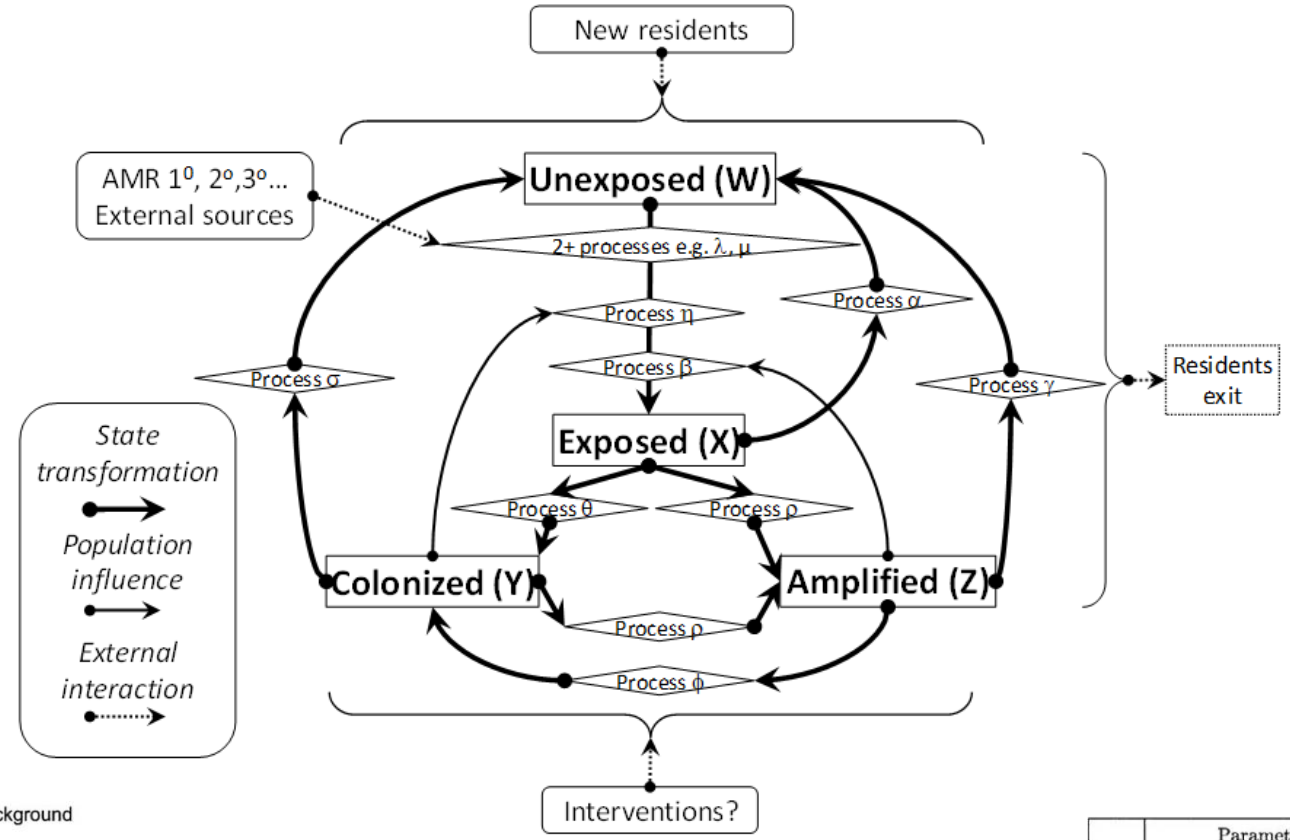
II. Products

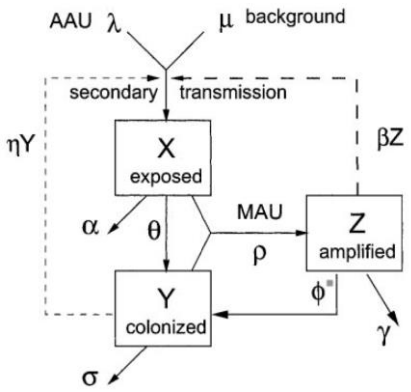
- Understanding system and its likely evolution
- Research & resource allocation priorities
- Revised monitoring
- Risk definition
- ISO compliant Risk management
- Quantifying bloom potential
- Decision support
- Costs v. benefits
- Auditable managem and decision basis



Smith et al. 2002
(environment<>RACF interaction)

Timeseries/ dynamic Modelling used Reed Frost and Binomial formulation to **incorporate stochasticity**





SMITH, D. L., HARRIS, A. D., JOHNSON, J. A., SILBERGELD, E. K. & MORRIS, J. G. 2002. Animal antibiotic use has an early but important impact on the emergence of antibiotic resistance in human commensal bacteria. Proceedings of the National Academy of Sciences, 99, 6434-6439.

	Parameter	Estimate
λ, μ	Rates of Exposure	$10^{-3}, 10^{-6}$
α	Transient Loss Rate	0.1 per day
θ	Colonization Rate	0.001 per day
σ	Colonized Loss Rate	0.003 per day
γ	Amplified Loss Rate	0.007 per day
ϕ	Recolonization Rate	0.003 per day
ρ	Prescription Rate	0.003 per day
η	Contact Rate, colonized	10^{-5} per day
β	Contact Rate, Amplified	0.5 per day

How good was the predictive power?

total AMR and Multiply drug resistant E. coli simulated sets (n=1000)

```
For AMR_TF:
----------

Confusion:
  ...Predicted..
   FALSE    TRUE     Actual
  ------  ------    ------
     510      74     FALSE
     210     206     TRUE

Error rate = 28.4%

Scoring Rule Results:
  Logarithmic loss = 0.5391
  Quadratic loss   = 0.3745
  Spherical payoff = 0.7855
```

```
For EcoliMDR_TF:
---------------

Confusion:
  ...Predicted..
   FALSE    TRUE     Actual
  ------  ------    ------
     718      54     FALSE
     138      90     TRUE

Error rate = 19.2%

Scoring Rule Results:
  Logarithmic loss = 0.4003
  Quadratic loss   = 0.2649
  Spherical payoff = 0.8515
```

Variables discarded

- Initial Screening – multiple R clusters > 0.1 for AMR bacteria incidence variables versus candidate drivers
- Modelling: Identify best explanatory variables and compare
- For multiple regression (numeric AMR data) use Backward Regression repeatedly reducing P value and maximising r^2
- For logistic regression (T/F AMR data variants) remove high P values to reduce variable set and maximise r^2
- For Bayes Nets mutual information < 1% after semi-naive BN

Bayes v. Regression 3. –
Multiply resistant Staphylococcus aureus incidence

Bayes Net (% Variance reduction)		Backwards Regression (coefficient P.)	
RTI	**3.36**	**RTI**	**0.0098**
Antimicrobial_n	2.71	Cefalexin	0.0920
Mobility_rating	2.64	Systemic_TF	0.1911
Facility	2.51	Systemic_duration	0.1983
Systemic_duration	1.97		
Systemic_total	1.75		
Systemic_TF	1.55	(overall P. = 0.0056 and R^2 = 0.13)	
UTI	1.17		

('Facility' variable not suited to regression but fine for BNs)