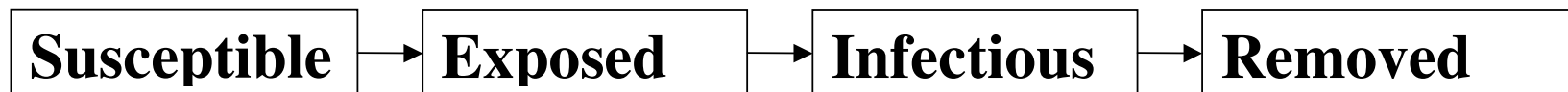


Statistical inference for mean contact rate in spatially structured SEIR models

Zsolt Lang, Jenő Reiczigel

Szent István University, Faculty of Veterinary Science,
Department of Biomathematics and Informatics, Budapest



We focus on the **Susceptible**→**Infectious** transmission:

Deterministic: $\dot{I}(t) = \beta \cdot S(t) \cdot I(t) + \dots$

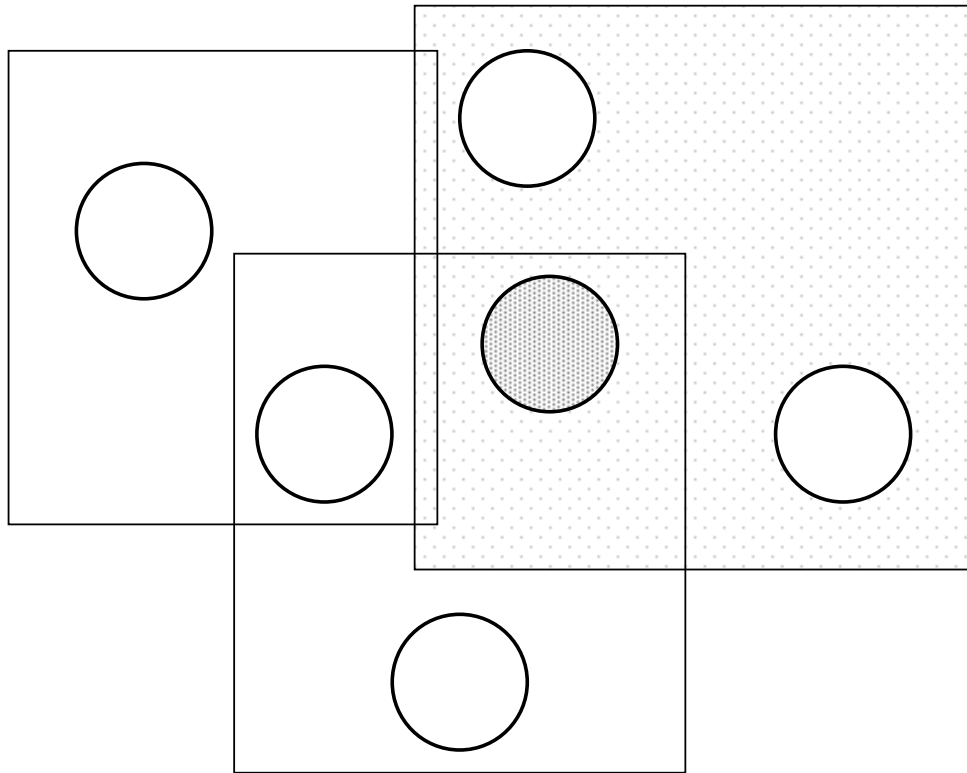
Stochastic: $P(I \rightarrow I + 1) = \beta \cdot S(t) \cdot I(t) + \dots$

β transmission rate

$S(t) \cdot I(t)$ maximum of possible contacts

Individuals do not make contacts with others living far away from them.

Spatial structure: locations and their vicinities



- Locations form a population
- Vicinities may be overlapping
- Clustered data

Illustrative example

Location	<i>S</i>	<i>I</i>	<i>O</i>	<i>PI</i>
1.	1	2	2	2/4
2.	2	1	3	1/4
3.	9	3	1	3/4
Average	4	2	2	2/4

All imaginable contacts

$$\sum S \cdot (I + O) = 12 \cdot 4 = 48$$

Spatially realizable contacts

$$\sum S \cdot I = 1 \cdot 2 + 2 \cdot 1 + 9 \cdot 3 = 31$$

$$CPIS = 31/48 \approx 0.65$$

- Contacts between infectious and susceptible individuals are reduced to 65% due to spatial structure
- Raw average of *PI* is 50%, the positive association of *S* and *I* accounts for the 15% increase in *CPIS*

Randomly selected location

S susceptible individuals in the location

I infectious individuals in the vicinity

O infectious individuals outside the vicinity

Local proportion of infectious individuals

$$PI = \frac{I}{I + O}$$

Only this proportion of infectious individuals is assumed to have contacts with the local susceptibles.

In the SEIR model replace

$$\beta \cdot S(t) \cdot I(t)$$

with

$$\beta \cdot \left(\frac{E(S \cdot PI)}{E(S)} \right) \cdot S(t) \cdot I(t)$$

CPIS

Contact proportion between infectious and susceptible individuals, averaged over all susceptibles.

The **contact proportion of infectious and susceptible individuals** depends on the raw average of PI over locations and the spatial correlation and heterogeneity:

$$CPIS = \frac{E(S \cdot PI)}{E(S)} = E(PI) + \frac{Corr(S, PI) \cdot \sigma(S) \cdot \sigma(PI)}{E(S)}$$

- Positive spatial correlation of S and PI increases, negative correlation decreases $CPIS$
- Large spatial variation of both S and PI strengthens the impact of spatial correlation on $CPIS$

Statistical inference

Assumption: *CPIS* does not vary over time

Random cross-sectional sample (with replacement)

$$S_i, PI_i \quad i = 1, 2, \dots, n$$

Estimating formula

$$\hat{C} = \frac{\sum S_i \cdot PI_i}{\sum S_i}$$

Let $\underline{S} = (S_1, \dots, S_n)^T$.

Variance decomposition:

$$\text{Var}(\hat{C}) = \text{Var}(E(\hat{C} | \underline{S})) + E(\text{Var}(\hat{C} | \underline{S}))$$

Theorem 1:

$$E(\hat{C} | \underline{S}) = \frac{\sum S_i E(PI_i | S_i)}{\sum S_i}$$

If the nonlinear regression function

$$f(S) = E(PI | S)$$

is known from a preliminary analysis, then

$$\hat{C}_f = E(\hat{C} | \underline{S}) = \frac{\sum S_i f(S_i)}{\sum S_i}$$

is a better estimate than \hat{C} , because it has the same bias, but its variance is smaller.

Moreover, \hat{C}_f depends only on the frequencies of susceptible individuals.

Theorem 2:

If f is monotone increasing (decreasing), then \hat{C}_f has a negative (positive) bias. (Reiczigel et al. 2005)

The greater is the skewness of the distribution of S , the larger is the bias of \hat{C}_f .

Large sample properties of \hat{C}_f :

- asymptotically normally distributed
- the magnitude of its bias is $\frac{1}{n}$

(Cochran 1963)

Bootstrap tests and confidence intervals are also referred to in (Reiczigel et al. 2005).