Some notation

- For each of the *m* regions on the map:
- y_i or n_i the count of disease in the i^{th} region
- e_i is the expected count in the i^{th} region
- θ_i is the relative risk in the i^{th} region
- The SMR is just $smr_i = y_i / e_i$
- This is just an estimate of θ_i

Standardization

- Direct: conversion of counts to rates
- Indirect: compute the expected number of cases based on a reference population
- Indirect used here extensively as we have rare diseases and sparse data
- Computation: must know reference population rate
  - e.g. counties and statewide rate (R):

$$R = \sum_{counties} y_i \Big/ \sum_{counties} p_i, \text{ where } p_i \text{ is the county population}$$

  - Then:

$$e_i = p_i.R = y_T.(p_i / p_T)$$

SMR problems

- Notoriously unstable
- Small expected count can lead to large SMRs
- Zero counts aren't differentiated
- The SMR is ***just the data!***

Smoothing for risk estimation

- Modern approaches to relative risk estimation rely on smoothing methods
- These methods often involve additional assumptions or model components
- Here we will examine only one approach: ***Bayesian modeling***