Outline

- Hierarchy of Information Levels
- Final-size models
- Survival models
- Likelihood models
- Bayesian models
- Summary

Basic Setting for Household Studies

- A community of households. May consider neighborhoods.
- Infectious forces.
  - Community at large: zoonotic source, infectious visitors.
  - Within-household transmission.
  - Between-household transmission.
- Symptom diary, e.g., headache, sore throat, fever.
- Lab-confirmation:
  - Viral culture for nasal/throat swabs, often triggered by symptom onset.
  - HI titers: baseline and the end of study.
- Intervention implemented, e.g., vaccine vs. placebo.

Hierarchy of Information Levels

- Consecutive occurence of infections is a counting process, observed at different information levels (Rhodes, Halloran and Longini, JRSS B, 1996)
  - How many infections have occurred in (0, T]. Final value models (Longini et al, 1982; Addy et al, 1991).
  - Times at which infection or symptom onset occurs. Survival model (Longini and Halloran, 1996).
  - Who contacts whom and/or who infects whom. Discrete-time likelihood models (Rampey et al, 1992; Yang et al, 2006).
    - Sometimes difficult to obtain.
    - Clustering pattern is the bottom line.

Final Size Model

- Longini and Koopman (Biometrics, 1982)
  - B: Probability of escaping infection from external source during epidemic.
  - Q: Probability of escaping infection from an infectious household member during epidemic.
  - m_{jk}: probability that j out of k household members are infected.
    - Household with a single person: $m_{01} = B$ and $m_{11} = 1 - B$
    - Household with two members:
      - $m_{02} = B^2$
      - $m_{12} = 2(1 - B)BQ$
      - $m_{22} = 1 - m_{02} - m_{12} = 2(1 - B)(1 - Q)B + (1 - B)^2$
    - In general, $m_{jk} = \binom{k}{j} m_{jj} B^{k-j} Q^{j(k-j)}$ and $m_{jj} = 1 - \sum_{l<j} m_{lj}$.

- Maximum likelihood estimation
  - Likelihood: $L(B, Q) = \prod_{k,j} m_{jk}^{a_{jk}}$, where a_{jk} is the frequency of households corresponding to m_{jk}.
  - Score function:

$$\frac{\partial \ln L}{\partial B} = \sum_{k,j} a_{jk} \Big\{ \frac{1}{m_{jj}} \big(\frac{\partial m_{jj}}{\partial B} \big) + \frac{k-j}{B} \Big\}.$$

  - Fisher's information:

$$-E\Big(\frac{\partial^2 \ln L}{\partial B^2} \Big) = \sum_{k,j} n_k m_{jk} \Big\{ \frac{1}{m_{jj}^2} \big(\frac{\partial m_{jj}}{\partial B} \big)^2 - \frac{1}{m_{jj}^2} \frac{\partial^2 m_{jj}}{\partial B^2} + \frac{k-j}{B^2} \Big\}.$$

- $*$ Rough estimates for starting point

$$\frac{a_{0k}}{n_k} = \hat{m}_{0k} = \hat{B}_k^k \Rightarrow \hat{B}_k = \big(\frac{a_{0k}}{n_k}\big)^{1/k} \Rightarrow \hat{B} = \frac{1}{n}\sum_k n_k \hat{B}_k$$

$$\frac{a_{1k}}{n_k} = \hat{m}_{1k} = k(1-\hat{B})\hat{B}^{k-1}\hat{Q}^{k-1}$$

$$\hat{Q}^{\hat{\phi}}\hat{B} \approx 1-\hat{\theta} \Longrightarrow \hat{Q} \approx \big(\frac{1-\hat{\theta}}{\hat{B}}\big)^{1/\hat{\phi}}$$

where $\hat{\phi} = \frac{\sum_{k,j} j a_{jk}}{n}$ and $\hat{\theta} = \frac{\sum_{k,j}(j a_{jk}/k)}{n}$.

- Inter-group mixing (Addy, Longini and Haber, Biometrics, 1991).

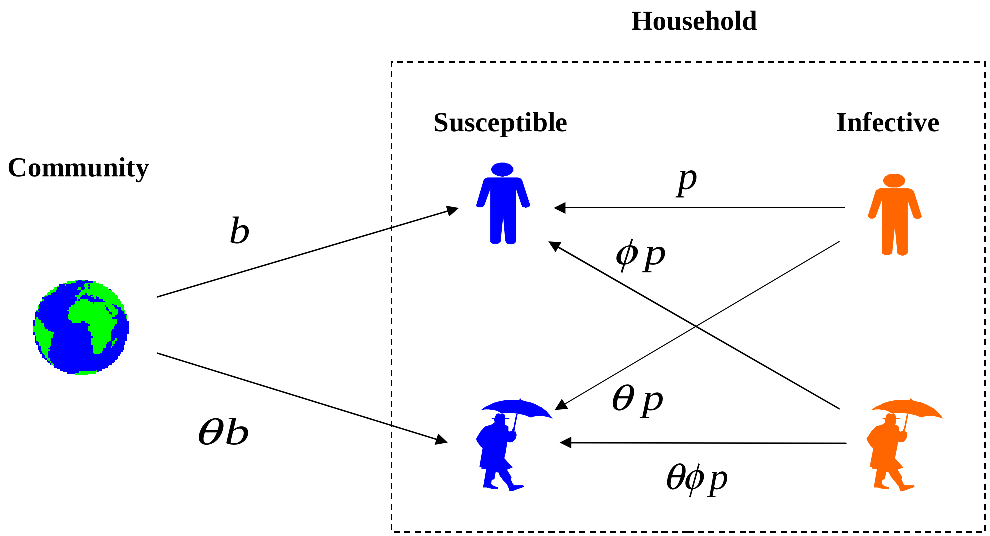
Frailty Hazard Model

- Longini and Halloran (Applied Stat, 1996)
  - α_v: proportion of full immunity in group v ($1 =$ vaccine, $0 =$ control).
    - If $\alpha_1 > \alpha_0$, "all-or-none"effect.
  - θ: reduction rate in susceptibility for the $1 - \alpha_1$ of vaccinated population, "leaky"effect.
  - Frailty (random) hazard
    - $\Pr(Z_v = 0) = \alpha_v$
    - $Z_v | Z_v > 0 \sim f_v(\text{mean} = 1, \text{variance} = \delta_v)$
    - Hazard function: $\lambda_v(t) = Z_v \theta^v c \pi p(t)$.
    - Survival function: $S_v(t) = E_{Z_v}\Big[\exp\big\{-Z_v \int_0^t \lambda_v(\tau) d\tau\big\}\Big]$

- $VE = 1 - \frac{E\left[\lambda_1(t)\right]}{E\left[\lambda_0(t)\right]} = 1 - \frac{(1-\alpha_1)\theta c\pi p(t)}{(1-\alpha_0)c\pi p(t)} = 1 - \frac{(1-\alpha_1)}{(1-\alpha_0)}\theta.$
- For grouped survival data with k intervals, $p(t) = \prod_{i=1}^{k} p_i^{I(t_{i-1} \le t < t_i)}$.
- r_{iv}: number of subjects at risk at the beginning of iterval $[t_{i-1}, t_i)$.
- m_{iv}: number of subjects infected in iterval $[t_{i-1}, t_i)$.
- Likelihood function

$$L = \prod_{i=1}^{k} \prod_{v=0}^{1} \left\{ \frac{S_v(t_i)}{S_v(t_{i-1})} \right\}^{r_{iv} - m_{iv}} \left\{ 1 - \frac{S_v(t_i)}{S_v(t_{i-1})} \right\}^{m_{iv}}$$

Transmission Patterns and Parameters of Interest



Household

Susceptible

Infective

Community

b

θb

p

ϕp

θp

$\theta\phi p$

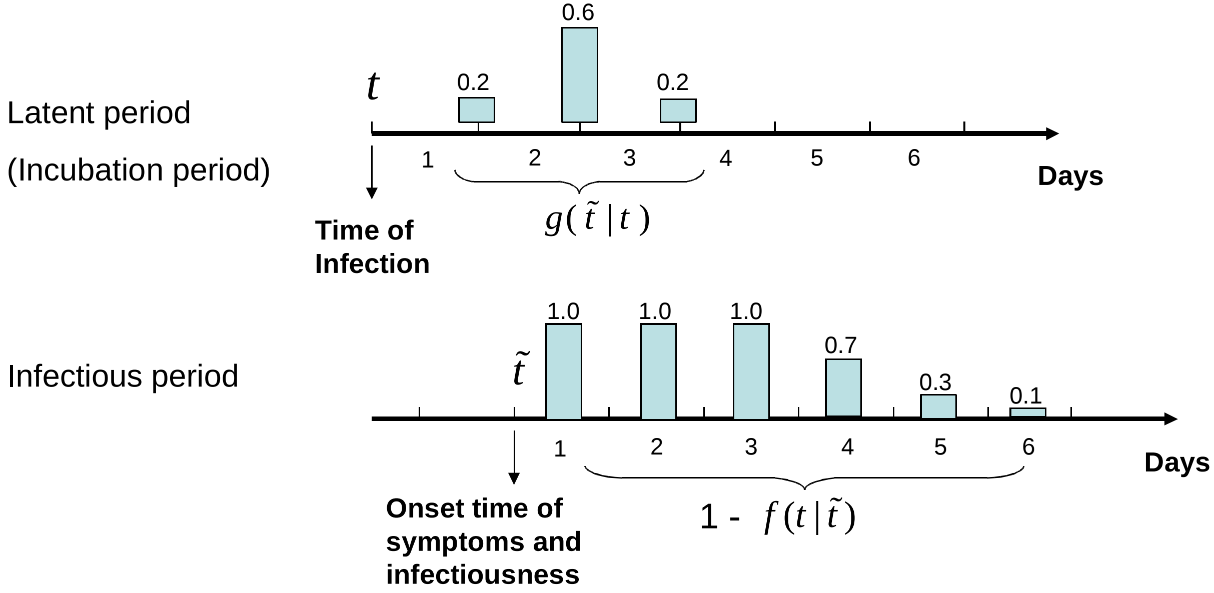
p: within-household pairwise daily transmission probability without treatment.

b: daily probability of infection by the community without treatment (CPI).

AVE$_S$ = $1-\theta$: Efficacy of the antiviral agent in reducing susceptibility.

AVE$_I$ = $1-\phi$: Efficacy of the antiviral agent in reducing infectiousness.

Natural Disease History of Influenza



$g(\tilde{t} \mid t)$: The probability of symptom onset on day $\tilde{t}$ given infection on day t.

$f(t \mid \tilde{t})$: Probability that the host is infective on day t given symptom onset on day $\tilde{t}$.

Likelihood Model for Symptomatic Infection

Yang, Longini & Halloran (Appl. Stat., 2006)

- Likelihood for a person-day

  Probability of pairwise transmission per daily contact:

$$p_{ji}(t) = \begin{cases} \theta^{r_i(t)}\phi^{r_j(t)} p f(t|\tilde{t}_j), & j \in H_i \\ \theta^{r_i(t)} b, & j = c. \end{cases}$$

  Define $D_i = H_i \cup c$. Probability of escaping infection on day t:

  $e_i(t) = \prod_{j \in D_i} \big(1 - p_{ji}(t)\big)$

  Probability of escaping infection up to day t:

  $Q_i(t) = \prod_{\tau=1}^{t} e_i(\tau)$

- Likelihood contributed by a single individual

  If subject i is known to be infected on day t, the probability is

  $$U_i(t) = [1 - e_i(t)]\, Q_i(t-1),$$

  Generally only symptom onset is observable

  $$L_i = \begin{cases} Q_i(T), & \text{if individual } i \text{ is not infected} \\ \sum_{t=\underline{t_i}}^{\overline{t_i}} g(\tilde{t}_i|t) U_i(t), & \text{otherwise} \end{cases}$$

  where $\underline{t_i} = \tilde{t}_i - l_{max}$, $\overline{t_i} = \tilde{t}_i - l_{min}$ and T is the last observation day for the epidemic.

- Selection bias in case-ascertained design: only households with infected members are followed.
  - Conditioning on the disease history (infection and symptom) up to the symptom onset day of the index case $\tilde{t}_{d_i}$.

$$L_i^m = \begin{cases} L_i, & \text{index case,} \\ \sum_{t=1}^{\tilde{t}_{d_i}} \Big\{ U_i(t) \Pr(\tilde{t}_i > \tilde{t}_{d_i} | t) \Big\} + Q_i(\tilde{t}_{d_i}), & \text{otherwise.} \end{cases}$$

  - Use the conditional likelihood $L_i^c = L_i / L_i^m$ for inference.

- Right-censoring: real-time analysis
  - No symptoms observed could mean either escape from infection or incubation period.
  - Calculate the marginal probability of observing no symptom onset up to day T:

$$L_i^m = Q_i(T - l_{min}) + \sum_{t=T-l_{max}+1}^{T-l_{min}} \Big\{(1 - e_i(t))Q_i(t-1)\Big\} \times \Pr(\tilde{t}_i > T|t)$$

- Assessing goodness of fit
  - The probability of symptom onset on day t for subject i is

$$\pi_i(t) = \sum_{\tau=t-l_{max}}^{t-l_{min}} \Big\{ \big(1 - e_i(\tau)\big) \prod_{s=t-l_{max}}^{\tau-1} e_i(s) \Big\} g(t|\tau).$$

  - Choose $0 = c_0 < c_1 < \ldots < c_m = 1$, then $\hat{n}_k = \sum_{c_{k-1}<\pi_i(t)<c_k} \pi_i(t)$ is the fitted count in level k. Let N_k be the total person-days and $\tilde{n}_k$ be the observed count in level k, then

$$\sum_{k=1}^{m} \frac{N_k(\tilde{n}_k - \hat{n}_k)^2}{\hat{n}_k(N_k - \hat{n}_k)} \sim \chi^2_{m-2}.$$

  - If $\hat{n}_k \ll N_k$ for all k, it is simplified to $\sum_{k=1}^{m} \frac{(\tilde{n}_k - \hat{n}_k)^2}{\hat{n}_k}$.

- Simulation study

  Population: a community composed of households of size two or larger with 1000 people is generated based on the age distribution and household sizes from the US Census 2000.

Table 1: Empirical distributions of the latent period and the infectious period (Elveback et al., 1976)

Latent Period		Infectious Period	
(days)	Com. Prob.	(days)	Cum. Prob.
1	0.2	3	0.3
2	0.8	4	0.7
3	1.0	5	0.9
		6	1.0

Table 2: Comparison of MLEs by randomization schemes and household follow-up schemes

Parameter‡		*Estimate*		*MonteCarlo standarderrors*		95%*CIcoverage* (%)§§	
		I§	*H*§	*I*§	*H*§	*I*§	*H*§
θ							
	Prospective	0.70	0.71	0.083	0.25	95.3	93.8
	Case-ascertained	0.70	0.71	0.083	0.26	96.1	94.3
ϕ							
	Prospective	0.20	0.24	0.045	0.16	94.6	91.5
	Case-ascertained	0.20	0.24	0.044	0.15	95.3	91.3

‡ True efficacy-related parameters are set to $\theta = 0.70$ and $\phi = 0.20$.

§ I, individual-level randomization; H, household-level randomization.

§§ The 95% CI is obtained as $\exp[\log(\hat{\lambda}) \pm 1.96 \times \mathrm{se}\{\log(\hat{\lambda})\}]$; $\lambda = \theta, \phi$.

Table 3: Two randomized multi-center trials of Oseltamivir, an influenza antiviral agent.

	Trial I (Welliver et al. 2001)	Trial II (Hayden et al. 2004)
Time of trial	1998-1999	2000-2001
Households	372	277
Population	1329	1110
Treatment for illness	None	Oseltamivir
Duration of medication		
Illness treatment	N/A	5 days
Prophylaxis	7 days	10 days
Follow up (symptom diary)	14 days	30 days
Infected/Exposed(index)	165/372	179/298
Infected/Exposed(susceptible)		
Control†	38/464	45/392
Oseltamivir	4/493	14/420

Table 4: Maximum likelihood estimates by age (1-17 vs 18+) for pooled oseltamivir trials conducted in 1998-1999 and 2000-2001, North America and Europe.

With Assumption $\psi = \theta\phi$	Parameter	MLE	95% C.I.
Yes	b_c[†]	0.0023	(0.0015, 0.0035)
	b_a	0.00055	(0.0003, 0.001)
	p_{cc}	0.038	(0.023, 0.063)
	p_{ca}	0.012	(0.007, 0.021)
	p_{ac}	0.018	(0.008, 0.040)
	p_{aa}	0.022	(0.014, 0.034)
	AVE_S	0.85	(0.52, 0.95)
	AVE_I	0.66	(-0.10, 0.89)
	AVE_T	0.95	(0.77, 0.99)
No	AVE_S	0.93	(0.50, 0.99)
	AVE_I	0.78	(-0.27, 0.96)
	AVE_T	0.87	(0.41, 0.97)
	SAR_{cc}[‡]	0.15	(0.074, 0.21)
	SAR_{ca}	0.049	(0.021, 0.075)
	SAR_{ac}	0.071	(0.014, 0.13)
	SAR_{aa}	0.086	(0.047, 0.12)

†, ‡ Subscription *c* denotes child (1-17), *a* denotes adult (18+), and *ca* denotes child-to-adult transmission.

‡ SAR_{vu} is based on the average 4.1 days of infectious period, i.e., $\text{SAR}_{vu} = 1 - (1 - p_{vu})^{4.1}$.

Table 5: Assessing goodness-of-fit of the likelihood model† for pooled oseltamivir trials conducted in 1998-1999 and 2000-2001, North America and Europe.

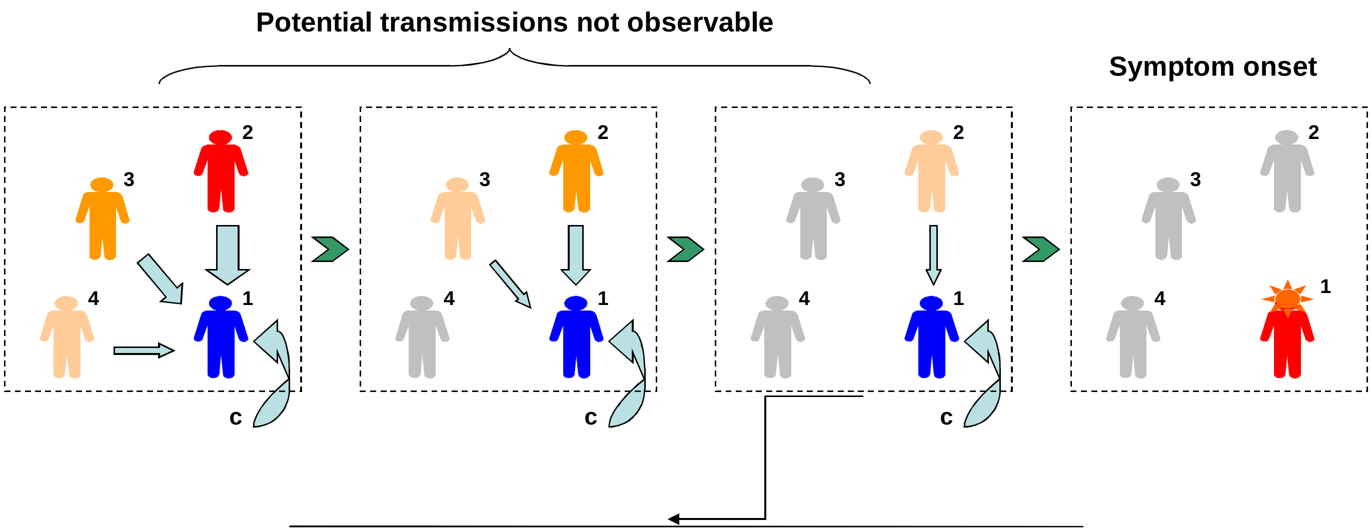
Risk Level	Total Person-days	Observed # of illness onsets	Predicted # of illness onsets
1	2084	0	0
2	1321	0	0
3	15878	8	9
4	1434	1	3
5	8165	19	22
6	933	3	3
7	935	5	4
8	1241	12	9
9	1084	17	18
10	894	25	27

† With assumption of $\psi = \theta\phi$.

Likelihood Model with Data Augmentation

Yang, Longini and Halloran (Comp. Stat. & Data Analysis, 2007)

- **Data to be augmented**
  - Pairwise transmission outcome $Y_{ji}(t)$ (1:transmission, 0:escape).
  - $Y_{ji}(t)$ is defined only if $Y_{ji}(\tau) = 0$ for all $\tau < t$.
  - $Y_{ji}(t)$ is not observed when j is infectious and $\underline{t_i} \le t \le \overline{t_i}$.
  - $Y_{ji}(t)$ is independent of $Y_{ki}(t)$ for the same day t.
  - More convenient to work with $Z_{ji}(t) = Y_{ji}(t) \prod_{k \in D_i, \tau < t} \big(1 - Y_{ki}(\tau)\big)$ and $\bar{Z}_{ji}(t) = \big(1 - Y_{ji}(t)\big) \prod_{k \in D_i, \tau < t} \big(1 - Y_{ki}(\tau)\big)$.



Potential transmissions not observable

Symptom onset

Exposure i	j	Outcome (1: transmission, 0:escape)	Expected Frequency
1	2	0	$\Pr[\, \bar{Z}_{21}(\tilde{t}_1 - 1) = 1 \mid I_1(\tilde{t}_1)\,]$
		1	$\Pr[\, Z_{21}(\tilde{t}_1 - 1) = 1 \mid I_1(\tilde{t}_1)\,]$
1	c	0	$\Pr[\, \bar{Z}_{c1}(\tilde{t}_1 - 1) = 1 \mid I_1(\tilde{t}_1)\,]$
		1	$\Pr[\, Z_{c1}(\tilde{t}_1 - 1) = 1 \mid I_1(\tilde{t}_1)\,]$

- **The likelihood of the augmented data**

$$
\begin{aligned}
&L_i(b, p, \theta, \phi | \tilde{t}_j, Z_{ji}(t), \bar{Z}_{ji}(t), j \in D_i, t \leq T) \\
=&\prod_{t=1}^{T}\Big\{g(\tilde{t}_i|t)^{max_{j\in D_i} Z_{ji}(t)} \prod_{j\in D_i} (p_{ji}(t))^{Z_{ji}(t)}\big(1 - p_{ji}(t)\big)^{\bar{Z}_{ji}(t)}\Big\},
\end{aligned}
$$

where $max_{j\in D_i} Z_{ji}(t)$ indicates if $Z_{ji}(t) = 1$ for any j on day t. The log-likelihood is

$$
\begin{aligned}
&\log(L_i(b, p, \theta, \phi | \tilde{t}_j, Z_{ji}(t), \bar{Z}_{ji}(t), j \in D_i, t \leq T)) \\
\propto&\sum_{t=1}^{T} \sum_{j\in D_i} \Big\{Z_{ji}(t) \log(p_{ji}(t)) + \bar{Z}_{ji}(t) \log(1 - p_{ji}(t))\Big\},
\end{aligned}
$$

- **The E-M algorithm** Define the events
  - $S_i(t)$: i has symptom onset on day t.
  - $I_i(t)$: i is infected on day t.
  - $I_{ji}(t)$: j infects i on day t.

  whose probabilities are given by

$$
\begin{aligned}
\Pr[I_{ji}(t)] &= \hat{Q}_i(t-1)\hat{p}_{ji}(t) \\
\Pr[I_i(t)] &= \hat{Q}_i(t-1)\big\{1-\hat{e}_i(t)\big\}, \\
\Pr\big[S_i(\tilde{t}_i)\big] &= \sum_{\tau=\underline{t_i}}^{\bar{t_i}} g(\tilde{t}_i|\tau) \times \Pr\big[I_i(\tau)\big],
\end{aligned}
$$

The conditional distributions of $Z_{ji}(t)$ and $\bar{Z}_{ji}(t)$ are

$$\Pr(Z_{ji}(t) = 1 | b, p, \theta, \phi, \tilde{t}_i) = \begin{cases} \frac{\Pr\left[I_{ji}(t)\right]}{\Pr\left[S_i(\tilde{t}_i)\right]} \times g(\tilde{t}_i | t), & \underline{t_i} \leq t < \bar{t_i} \\ 0, & \text{otherwise} \end{cases}$$

and

$$\Pr(\bar{Z}_{ji}(t) = 1 | b, p, \theta, \phi, \tilde{t}_i)$$

$$= \begin{cases} \frac{g(\tilde{t}_i|t) \times \left\{\Pr\left[I_i(t)\right] - \Pr\left[I_{ji}(t)\right]\right\}}{\Pr\left[S_i(\tilde{t}_i)\right]} + \sum_{\tau=t+1}^{\bar{t_i}} \frac{g(\tilde{t}_i|\tau) \times \Pr\left[I_i(\tau)\right]}{\Pr\left[S_i(\tilde{t}_i)\right]}, & \underline{t_i} \leq t < \bar{t_i} \\ 1, & t < \underline{t_i} \\ 0, & \text{otherwise} \end{cases}$$

- **Variance estimation**

  Let $\boldsymbol{Z} = \{Z_{ji}(t), \bar{Z}_{ji}(t)\}$, $\tilde{\boldsymbol{t}} = \{\tilde{t}_i\}$, and $\boldsymbol{\lambda} = \{b, p, \theta, \phi\}$. Louis' method states that

$$\begin{aligned}\frac{\partial^2 \log(L(\boldsymbol{\lambda}|\tilde{\boldsymbol{t}}))}{\partial \boldsymbol{\lambda}^2} &= \mathrm{E}_{\boldsymbol{Z}|\tilde{\boldsymbol{t}},\boldsymbol{\lambda}}\Big\{-\frac{\partial^2 \log(L(\boldsymbol{\lambda}|\tilde{\boldsymbol{t}},\boldsymbol{Z}))}{\partial \boldsymbol{\lambda}^2}\Big\} \\ &\quad + \mathrm{VAR}_{\boldsymbol{Z}|\tilde{\boldsymbol{t}},\boldsymbol{\lambda}}\Big\{-\frac{\partial \log(L(\boldsymbol{\lambda}|\tilde{\boldsymbol{t}},\boldsymbol{Z}))}{\partial \boldsymbol{\lambda}}\Big\}\end{aligned}$$

Table 6: Two randomized multi-center trials of zanamivir, an influenza antiviral agent

	Hayden et al., 2000	Monto et al., 2002
Time of trial	Oct. 1998 - Apr. 1999	Jun. 2000 - Apr. 2001
Households	336	484
Population	1186	1770
Index case randomization	Yes	No
Duration of medication		
Index case	5 days	N/A
Contact	10 days	10 days
Follow up (symptom diary)	14 days	14 days
Infected† /Symptomatic(index)	164/336	281/484
Infected† /Exposed(contacts)		
Control	52/435	76/626
Zanamivir	17/415	27/660

Numbers may slightly differ from references due to different criteria of data inclusion for analysis.

† Laboratory-confirmed infections with clinical symptoms

Table 7: Estimates of efficacies and transmission probabilities by age (1-17 vs. 18+) for pooled zanamivir trials conducted in 1998-1999 and 2000-2001.

	IRLS		MLE		
Parameter	Point Estimate	SD	Point Estimate	SD	95% CI
b_c †	0.0024	0.00052	0.0028	0.00063	(0.0017, 0.0042)
b_a	0.00086	0.00030	0.0010	0.00039	(0.00045, 0.0021)
$p_{cc}^{\dagger}$	0.040	0.0074	0.040	0.0077	(0.027, 0.057)
p_{ca}	0.028	0.0045	0.029	0.0048	(0.021, 0.040)
p_{ac}	0.023	0.0071	0.020	0.0071	(0.009, 0.037)
p_{aa}	0.040	0.011	0.032	0.011	(0.016, 0.058)
AVE_S	0.68	0.086	0.75	0.072	(0.56, 0.86)
AVE_I	0.24	0.38	0.23	0.44	(-1.33, 0.75)
AVE_T			0.81	0.094	(0.50, 0.93)

† Subscript c denotes child (1-17), a denotes adult (18+), and ca denotes child-to-adult transmission.

A Bayesian Model for Symptomatic Infection

Cauchemez et al. (Stat in Med., 2004)

- Assumption: infectious period (v_i, ψ_i) follows Gamma(μ, σ).
- Observation level $P(Y|v,\psi) = \prod_{i \in I} \mathbf{1}\{Z_i - 3 < v_i < Z_i \text{ and } v_i < \psi_i\}$
- Transmission level

$$
\begin{aligned}
\lambda_i(t) &= \alpha_i + \epsilon_i \sum_{j \in I(t)} \frac{\beta_j}{n} \\
P(v,\psi|\theta) &= \prod_{i \in I} d_{\mu,\sigma}(\psi_i - v_i) \prod_{i \in I - \{1\}} \lambda_i(v_i) e^{-\int_{v_1}^{v_i} \lambda_i(t)dt} \prod_{j \in S} e^{-\int_{v_1}^{15} \lambda_j(t)dt} \\
\mathrm{SAR}_{i \to j}(n) &= 1 - \int_0^\infty \exp\Big(-\epsilon_j \frac{\beta_i}{n} t\Big) d_{\mu,\sigma}(t)dt
\end{aligned}
$$

- Prior level

A Bayesian Model for Asymptomatic Infections

Yang, Halloran & Longini (Biostatistics, 2009)

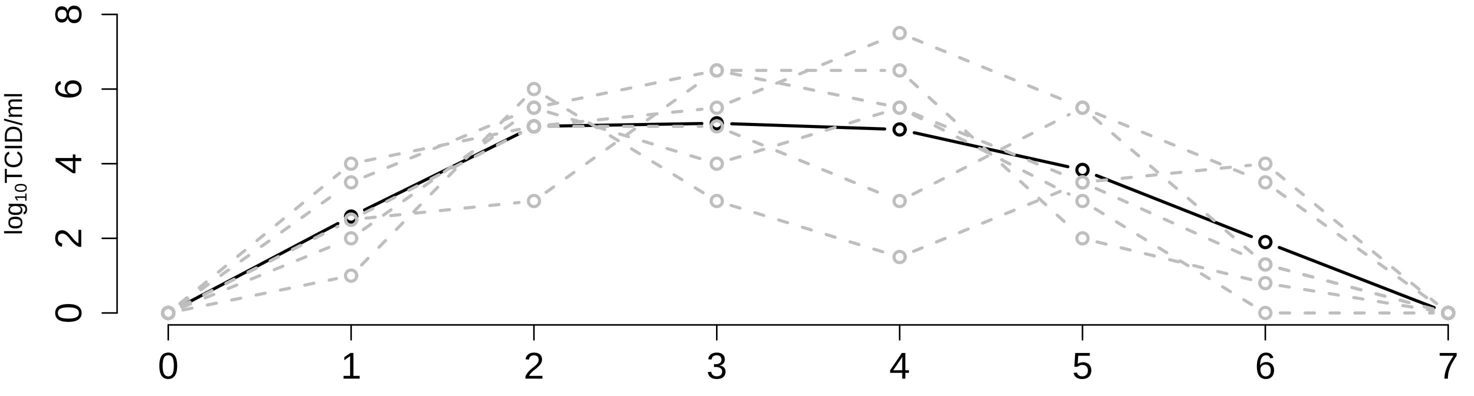
- Observed data
  - H households, with household h of size n_h.
  - Enrollment day (ascertainment of index cases) as day 1.
  - Follow-up period: $\{1, 2, \ldots, \overline{T_h}\}$.
  - ILI onset dates: $\{\tilde{t}_{hi} : i = 1, \ldots, n_h, h = 1, \ldots, H\}$.
  - Laboratory test results $\boldsymbol{y}_{hi}$.
- Latent variable
  - Infection dates: $\{\hat{t}_{hi} : i = 1, \ldots, n_h, h = 1, \ldots, H\}$.

- Modeling viral transmission
  - Exposure to constant risk γ_0 from community, starting from day $\underline{T_h} < 1$.
  - Exposure to time-varying baseline risk $\gamma_1 f(\frac{t-\hat{t}_{hj}}{\Delta}|a,b)$ from infected household member j during the infectious period $\hat{t}_{hj} \le t \le \hat{t}_{hj} + \Delta$.
    - Δ: duration of infectious period, a known constant.
    - $f(\cdot|a,b)$: beta density with parameters a and b.
    - γ_1: average risk over the infectious period, because $\gamma_1 = \frac{1}{\Delta}\int_{\hat{t}_{hi}}^{\hat{t}_{hi}+\Delta} \gamma_1 f(\frac{t-\hat{t}_{hi}}{\Delta}|a,b)dt$

- Risk adjusted for covariates:

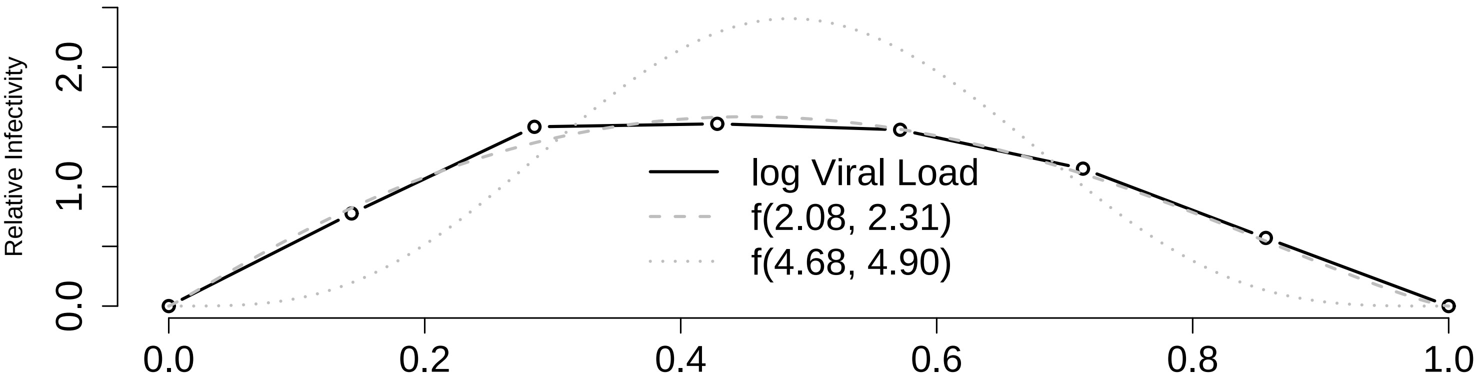
$$\lambda_{h,j\to i}(t) = \begin{cases} \gamma_1 f(\frac{t-\hat{t}_{hj}}{\Delta}|a,b)\exp\{\boldsymbol{\beta}'_S\boldsymbol{x}_{hi}(t)+\boldsymbol{\beta}'_I\boldsymbol{x}_{hj}(t)\}, & j>0 \text{ and } j\neq i, \\ \gamma_0\exp\{\boldsymbol{\beta}'_S\boldsymbol{x}_{hi}(t)\}, & j=0. \end{cases}$$

- Total risk: $\lambda_{hi}(t) = \sum_{j\neq i}\lambda_{h,j\to i}(t)$.
- Daily probability of escaping influenza infection: $q_{hi}(t) = \exp\left\{-\int_{t-1}^{t}\lambda_{hi}(\tau)d\tau\right\}$.



$\log_{10}$TCID/ml
0
2
4
6
8
0
1
2
3
4
5
6
7

(a)



Relative Infectivity
0.0
1.0
2.0
0.0
0.2
0.4
0.6
0.8
1.0
log Viral Load
f(2.08, 2.31)
f(4.68, 4.90)

(b)

- Modeling pathogenicity (probability of developing ILI)
  - Traditional assumption: the probability of ILI given infection, $\xi_{hi}(t)$, depends only on antiviral treatment status $r_{hi}(t)$ of susceptible i on day t.
    - $\xi_{hi}(t) = \eta_0^{1-r_{hi}(t)} \eta_1^{r_{hi}(t)}$, where η_u is the probability of developing ILI for $r_{hi}(t) = u$, $u = 0, 1$.
    - Traditional antiviral efficacy in reducing pathogenicity is defined as $\mathrm{AVE}_P = 1 - \eta_1/\eta_0$.

- Our assumption: $\xi_{hi}(t)$ depends on antiviral treatment status of the susceptible and all infectives in the household.
  - Consider a single infective j.
    - $\text{logit}\Big(\xi_{hi}(t)\Big) = \alpha_{00} + \alpha_{10} r_{hj}(t) + \alpha_{01} r_{hi}(t)$.
    - Let η_{uv} be the value of $\xi_{hi}(t)$ for $r_{hj}(t) = u$ and $r_{hi}(t) = v$, then, $\eta_{uv} = \text{logit}^{-1}(\alpha_{00} + v\alpha_{01} + u\alpha_{10})$

- For multiple infectious sources, consider the average treatment status weighted by cumulative risks.

$$\begin{aligned}
\Lambda_{h,j\to i}(t) &= \int_{t-1}^{t} \lambda_{h,j\to i}(\tau)d\tau, \\
\Lambda_{h,i}(t) &= \int_{t-1}^{t} \lambda_{h,i}(\tau)d\tau, \\
\bar{r}_h(t) &= \sum_{j=0}^{n_h} \left\{ \frac{\Lambda_{h,j\to i}(t)}{\Lambda_{h,i}(t)} r_{hj}(t) \right\}, \\
\text{logit}\Big(\zeta_{hi}(t)\Big) &= \alpha_{00} + \alpha_{10}\bar{r}_h(t) + \alpha_{01} r_{hi}(t).
\end{aligned}$$

- $*$ Antiviral efficacies in reducing pathogenicity:

$$\begin{aligned}\mathrm{AVE}_{Sp} &= 1 - \frac{\eta_{01}}{\eta_{00}},\\ \mathrm{AVE}_{Ip} &= 1 - \frac{\eta_{10}}{\eta_{00}},\\ \mathrm{AVE}_{Tp} &= 1 - \frac{\eta_{11}}{\eta_{00}}.\end{aligned}$$

- $*$ Antiviral efficacies for symptomatic infection:

$$\begin{aligned}1 - \mathrm{AVE}_{Sd} &= (1 - \mathrm{AVE}_{Si})(1 - \mathrm{AVE}_{Sp}),\\ 1 - \mathrm{AVE}_{Id} &= (1 - \mathrm{AVE}_{Ii})(1 - \mathrm{AVE}_{Ip}),\\ 1 - \mathrm{AVE}_{Td} &= (1 - \mathrm{AVE}_{Ti})(1 - \mathrm{AVE}_{Tp}).\end{aligned}$$

- For the incubation period $\tilde{t}_{hi} - \hat{t}_{hi}$, assume a known distribution $\Pr(\tilde{t}_{hi}|\hat{t}_{hi})$.
- Joint probability of viral transmission and ILI:
  - Define $\boldsymbol{\omega} = (\gamma_0, \gamma_1, \boldsymbol{\beta}_S, \boldsymbol{\beta}_I, \alpha_{00}, \alpha_{01}, \alpha_{10}, a, b)$.

$$L_{hi}\Big(\hat{t}_{hi}, \tilde{t}_{hi} \mid \boldsymbol{\omega}, \{\hat{t}_{hj} : j \neq i\}\Big)$$

$$= \begin{cases} \prod_{t=\underline{T_h}}^{\overline{T_h}} q_{hi}(t), & \hat{t}_{hi} > \overline{T_h}, \\ \Big\{\prod_{t=\underline{T_h}}^{\hat{t}_{hi}-1} q_{hi}(t)\Big\}\Big\{1 - q_{hi}(\hat{t}_{hi})\Big\}\xi_{hi}(\hat{t}_{hi}) \Pr(\tilde{t}_{hi}|\hat{t}_{hi}), & \underline{T_h} \leq \hat{t}_{hi} \leq \overline{T_h}, \tilde{t}_{hi} < \infty, \\ \Big\{\prod_{t=\underline{T_h}}^{\hat{t}_{hi}-1} q_{hi}(t)\Big\}\Big\{1 - q_{hi}(\hat{t}_{hi})\Big\}\big(1 - \xi_{hi}(\hat{t}_{hi})\big), & \underline{T_h} \leq \hat{t}_{hi} \leq \overline{T_h}, \tilde{t}_{hi} = \infty, \end{cases}$$

- Adjustment for selection bias
  - Each household in the analysis has at least one infection.
  - All index cases are symptomatic.
  - Without adjustment, estimates of γ_0, α_{00}, α_{01} and α_{10} will be biased.
  - Adjustment: drop likelihood history up to (include) $\widetilde{T_h}$, the symptom onset day of the index case.

$$L_{hi}\Big(\hat{t}_{hi}, \tilde{t}_{hi} \mid \boldsymbol{\omega}, \{\hat{t}_{hj} : j \neq i\}\Big)$$

$$= \begin{cases} \prod_{t=\widetilde{T_h}+1}^{\overline{T_h}} q_{hi}(t), & \hat{t}_{hi} > \overline{T_h}, \\ \Big\{\prod_{t=\widetilde{T_h}+1}^{\hat{t}_{hi}-1} q_{hi}(t)\Big\}\Big\{1 - q_{hi}(\hat{t}_{hi})\Big\}\xi_{hi}(\hat{t}_{hi}) \Pr(\tilde{t}_{hi}|\hat{t}_{hi}), & \widetilde{T_h} < \hat{t}_{hi} \leq \overline{T_h}, \tilde{t}_{hi} < \infty, \\ \Big\{\prod_{t=\widetilde{T_h}+1}^{\hat{t}_{hi}-1} q_{hi}(t)\Big\}\Big\{1 - q_{hi}(\hat{t}_{hi})\Big\}\big(1 - \xi_{hi}(\hat{t}_{hi})\big), & \widetilde{T_h} < \hat{t}_{hi} \leq \overline{T_h}, \tilde{t}_{hi} = \infty, \\ \Pr(\tilde{t}_{hi}|\hat{t}_{hi}), & \hat{t}_{hi} \leq \widetilde{T_h}, \tilde{t}_{hi} < \infty, \end{cases}$$

- Full probability
  - $\pi(\boldsymbol{\omega})$: the joint prior distribution.
  - $\hat{\boldsymbol{t}} = \{\hat{t}_{hi} : i = 1, \ldots, n_h, h = 1, \ldots, H\}$.
  - $\tilde{\boldsymbol{t}} = \{\tilde{t}_{hi} : i = 1, \ldots, n_h, h = 1, \ldots, H\}$.
  - $\boldsymbol{y} = \{\boldsymbol{y}_{hi} : i = 1, \ldots, n_h, h = 1, \ldots, H\}$
  - $C(\boldsymbol{y}_{hi}|\hat{t}_{hi})$: indicate whether $\boldsymbol{y}_{hi}$ is compatible with $\hat{t}_{hi}$ (1:yes, 0:no)

$$\begin{aligned}&\Pr(\hat{\boldsymbol{t}}, \boldsymbol{\omega}|\boldsymbol{y}, \tilde{\boldsymbol{t}}) \propto \Pr(\boldsymbol{y}, \hat{\boldsymbol{t}}, \tilde{\boldsymbol{t}}, \boldsymbol{\omega})\\ =&\pi(\boldsymbol{\omega}) \times \prod_{h=1}^{H} \prod_{i=1}^{n_h} L_{hi}\Big(\hat{t}_{hi}, \tilde{t}_{hi}; |\, \boldsymbol{\omega}, \{\hat{t}_{hj} : j \neq i\}\Big) C(\boldsymbol{y}_{hi}|\hat{t}_{hi}).\end{aligned}$$

- MCMC sampling
  - For parameters to be estimated, use random-walk style Metropolis-Hastings' algorithm.
  - For unobserved infection times,
    - the set of candidate infection days:

$$\Omega_{hi} = \begin{cases} \{t : C(\boldsymbol{y}_{hi}|t) \times L_{hi}(t, \tilde{t}_{hi}|\cdot) \Pr(\tilde{t}_{hi}|t) > 0\}, & \text{symptomatic,} \\ \{t : C(\boldsymbol{y}_{hi}|t) \times L_{hi}(t, \tilde{t}_{hi}|\cdot) > 0\}, & \text{asymptomatic} \end{cases}$$

    - Sample $\hat{t}_{hi}$ from

$$\Pr(\hat{t}_{hi} = t|\hat{t}_{hi} \in \Omega_{hi}, \cdot)$$

$$= \frac{L_{hi}\Big(t, \tilde{t}_{hi} \mid \boldsymbol{\omega}, \{\hat{t}_{hj} : j \neq i\}\Big) \prod_{j \neq i} L_{hi}\Big(\hat{t}_{hj}, \tilde{t}_{hj} \mid \boldsymbol{\omega}, \{\hat{t}_{hk} : k \neq j\}\Big)}{\sum_{s \in \Omega_{hi}} L_{hi}\Big(s, \tilde{t}_{hi} \mid \boldsymbol{\omega}, \{\hat{t}_{hj} : j \neq i\}\Big) \prod_{j \neq i} L_{hj}\Big(\hat{t}_{hj}, \tilde{t}_{hj} \mid \boldsymbol{\omega}, \{\hat{t}_{hk} : k \neq j\}\Big)}.$$

- Data Analysis
  - $\lambda_{h,j\to i}(t) = {\theta_{Rx}}^{\mathrm{RX}_i} {\phi_{Rx}}^{\mathrm{RX}_j} {\theta_{Age}}^{\mathrm{AGE}_i} {\phi_{Age}}^{\mathrm{AGE}_j} \gamma_1 f(\frac{t-\hat{t}_{hj}}{\delta}|a,b)$.
  - Identification of clinical symptom onset
    - I: $\geq 37.8^\circ$C plus cough or nasal congestion.
    - II: $\geq 37.2^\circ$C plus
      any of (cough, nasal congestion, sore throat) and
      any of (headache, aches/pains, chills/sweats, fatigue).
  - Identification of candidate infection days
    - A positive swab on day t indicates $t-\delta \leq \hat{t}_{hi} \leq t-1$.
    - 4-fold increase in HI titers indicates $1 \leq \hat{t}_{hi} \leq \overline{T}_h$ given that the subject is susceptible at baseline.

Table 8: Comparison between Bayesian estimates and previous findings.

Parameter	Bayesian	Halloran et al. (2007)†	Yang et al. (2006)‡
γ_0	$0.00046_{\ (0.00006,0.0017)}$		$0.00055_{\ (0.0003,0.001)}$
γ_1	$0.019_{\ (0.0096,0.037)}$		$0.022_{\ (0.014,0.034)}$
η_{00}	$0.50_{\ (0.33,0.67)}$	II: $0.57_{\ (0.44,0.69)}$	
η_{01}	$0.29_{\ (0.097,0.57)}$	II: $0.12_{\ (0.05,0.28)}$	
η_{10}	$0.082_{\ (0.017,0.22)}$		
AVE_{Si}	$0.62_{\ (0.39,0.77)}$	$\left(\begin{array}{l}\text{I} : 0.48_{\ (0.17,0.67)} \\ \text{II} : 0.64_{\ (0.36,0.80)}\end{array}\right)$	
AVE_{Ii}	$-0.18_{\ (-0.93,0.30)}$	$0.16_{\ (-0.33,0.46)}$	
AVE_{Sp}	$0.41_{\ (-0.28,0.81)}$	II: $0.79_{\ (0.45,0.92)}$	
AVE_{Ip}	$0.84_{\ (0.53,0.97)}$		
AVE_{Sd}	$0.77_{\ (0.45,0.93)}$	$\left(\begin{array}{l}\text{I} : 0.81_{\ (0.35,0.94)} \\ \text{II} : 0.91_{\ (0.64,0.98)}\end{array}\right)$	$0.85_{\ (0.52,0.95)}$
AVE_{Id}	$0.81_{\ (0.42,0.96)}$	$0.81_{\ (0.45,0.93)}$	$0.66_{\ (-0.10,0.89)}$
θ_{Age}	$1.06_{\ (0.64,1.91)}$		
ϕ_{Age}	$1.05_{\ (0.64,1.71)}$		
a	$4.68_{\ (1.44,16.86)}$		
b	$4.90_{\ (1.92,15.57)}$		
d	$3.88_{\ (3.03,4.48)}$		

Table 9: Bayesian estimates by different infectiousness of asymptomatic cases relative to symptomatic cases.

	Relative Infectiousness of Asymptomatic Infection				
	1.0	0.5	0.3	0.2	0.1
γ_0	0.00046 (0.00006,0.0017)	0.00063 (0.00009,0.0023)	0.0012	0.0025	0.0062 (0.0031,0.011)
γ_1	0.021 (0.011,0.038)	0.037 (0.019,0.067)	0.051	0.054	0.013 (0.0003,0.093)
η_{00}	0.49 (0.33,0.66)	0.48 (0.32,0.65)	0.45	0.38	0.29 (0.19,0.43)
η_{01}	0.30 (0.095,0.58)	0.31 (0.10,0.61)	0.32	0.34	0.35 (0.089,0.80)
η_{10}	0.080 (0.018,0.22)	0.077 (0.015,0.22)	0.076	0.079	0.063 (0.0,1.0)
AVE_{Si}	0.61 (0.36,0.78)	0.61 (0.35,0.77)	0.62	0.63	0.54 (-0.14,0.86)
AVE_{Ii}	-0.21 (-0.94,0.29)	-0.22 (-1.0,0.31)	-0.29	-0.29	0.28 (-12.62,0.98)
AVE_{Sp}	0.38 (-0.32,0.81)	0.35 (-0.46,0.79)	0.26	0.099	-0.18 (-2.06,0.71)
AVE_{Ip}	0.84 (0.53,0.96)	0.83 (0.49,0.97)	0.83	0.80	0.79 (-3.57,1.0)
AVE_{Sd}	0.77 (0.42,0.93)	0.75 (0.38,0.92)	0.72	0.68	0.50 (-0.46,0.86)
AVE_{Id}	0.80 (0.38,0.96)	0.80 (0.32,0.96)	0.78	0.73	0.81 (-5.13,1.0)
θ_{Age}	1.07 (0.64,1.87)	1.04 (0.64,1.74)	1.03	1.02	1.0 (0.59,1.77)
ϕ_{Age}	1.05 (0.64,1.69)	0.91 (0.56,1.55)	0.81	0.76	1.85 (0.25,38.50)

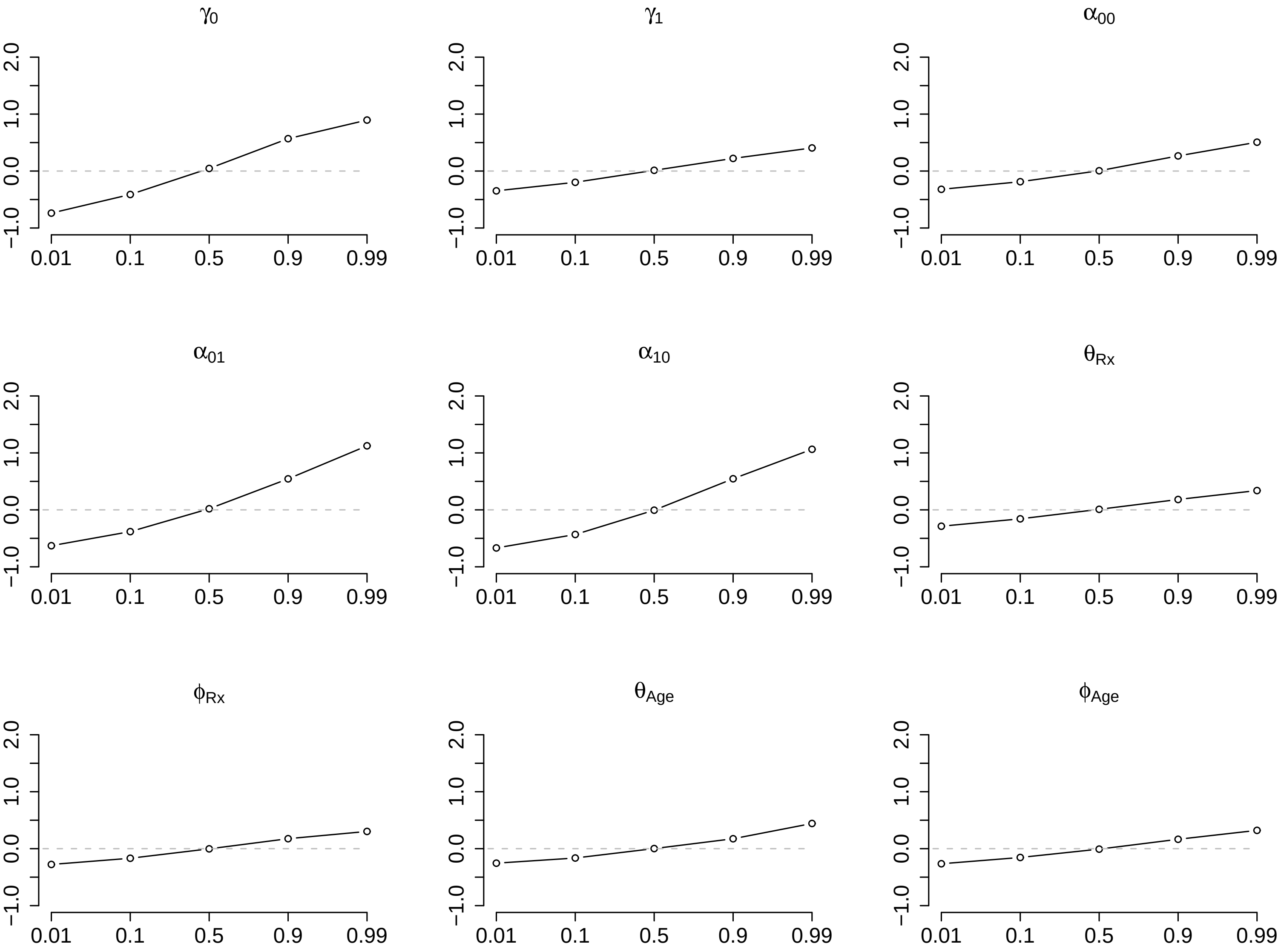


Figure 1: Sensitivity of the posterior median to the prior distribution for each parameter, with flat priors for parameters other than the focal one.

Discussion

- Methods designed for household studies may be generalizable to other cluster settings.
- Simpler methods may be more robust to model mis-specification, but may miss important information as well.
- Combining studies for meta analysis should be done carefully.
- Improve statistical inference via improving study design.
  - Maximize information for targeted efficacy measures.
  - individual level of randomization, including index cases.
  - Complete symptom diary.
  - Lab-tests at a higher frequency (e.g., the Hong Kong pilot NPI study).
- Post-randomization bias.