The Fractional Poisson: a Simple Dose-Response Model

Mike Messner
U. S. EPA
Office of Water
Messner.Michael@epa.gov

Image of Norovirus from RCSB
Outline

• Single-hit theory & models for microbial dose-response
• Norovirus & the beta-Poisson model
• The fractional Poisson model
• Other candidates for fractional Poisson
• What’s next?
Single-hit Theory

- A volunteer agrees to ingest a capsule containing a “known amount” of some microbial pathogen (could be virus, bacteria, protozoa, other “bugs”)
- The individual pathogen is “successful” if it overcomes any barriers and is able to initiate an infection in the host. As a result of the infection, lots and lots of newly-minted bugs are created and released to keep things going well for the bug (an poorly for the host).
- The single-hit idea is that, when any one bug is successful, the host is infected and what happens to the other bugs is not important. It only takes one.
- The host can’t be infected unless at least one bug is ingested.
Single-hit Models

• Two popular single-hit models:
  – Exponential
  – Beta-Poisson

• Exponential
  – Dose is Poisson.
  – Each bug succeeds with probability P, and independently of other bugs.

• Beta-Poisson
  – Same as exponential, but P varies from host-to-host as a beta random variable
Human Challenge Studies

• A series of human challenge studies have been conducted to identify norovirus infectivity.
• Human subjects ingest volumetrically-prepared doses.
  – Assumption: Particles are randomly dispersed, so actual number ingested is a Poisson random variable
  – Viruses were either aggregated (particle is many virions stuck together) or disaggregated (particle is single virion)
• Norovirus only binds to epithelial cells of subjects with positive ABH antigen secretor status (Se+). Se- subjects appear to be well-protected against infection.
  – Se- individuals do not possess a gene associated with a norovirus binding receptor.
  – We focus on Se+ subjects.
  – About 80% of people are Se+. 
Beta Distribution

- Two parameters (usually $\alpha$ and $\beta$)
- Defined between 0 and 1 (exclusive: 0 and 1 are excluded)
- Density function, $\text{dbeta}(r, \alpha, \beta)$, can have different shapes
  - Normal when $\alpha$ and $\beta$ are large
  - Lognormal when $\alpha$ is small and $\beta$ not
  - Bathtub when $\alpha$ and $\beta$ are both small
- Describes variation in host susceptibility
- Converges to exponential as $\alpha$ and $\beta$ become huge

- As $\alpha$ and $\beta$ become tiny, the probability mass is squeezed to 0 and 1.
  - 0 and 1 are really important.
  - But 0 and 1 are excluded.
  - Maybe some other model is needed.
Various Shapes for Beta-Poisson

- $\alpha = 1, \beta = 1$
- $\alpha = 0.5, \beta = 0.5$
- $\alpha = 5, \beta = 1$
- $\alpha = 1, \beta = 3$
- $\alpha = 2, \beta = 2$
- $\alpha = 2, \beta = 5$
- $\alpha = 5, \beta = 5$
Estimates are Based on Data

- Data from human challenge studies
- Se- subjects are immune, so we focus on the Se+ subjects. The virus is able to bind on cell surfaces and reproduce in only these subjects.
- At a particular dose, number infected is binomial random variable with parameters:
  - \( n = \text{number subjects} \)
  - \( p = \Pr\{\text{infection} \mid \text{dose}, \alpha, \beta\} \)
- Data includes \( k = \text{the number infected} \) (of the \( n \) subjects)
Beta-Poisson Infection Probability & Likelihood

- \( p = \Pr\{\text{infection} \mid \text{dose}, \alpha, \beta\} \) is given by

\[
\int_{0}^{1} dbeta(r, \alpha, \beta) \cdot \left( 1 - e^{-\text{dose} \cdot r} \right) \, dr
\]

- Number infected is \( k \) with probability (a.k.a. “likelihood”) = \( \text{dbinom}(k, n, p) \)
Data

• Four studies:
  – Teunis et al. 2008 (beta-Poisson model, 11 dose groups, some disaggregated)
  – Seitz et al. 2011 (1 disaggregated dose group)
  – Frenck et al. 2012 (1 aggregated dose groups)
  – Atmar et al. 2013 (4 aggregated dose group)

• A dose group is a set of subjects, each dosed at the same level (measured as genomic equivalent copies)
<table>
<thead>
<tr>
<th>Dose</th>
<th>Inoculum</th>
<th>Total Subjects</th>
<th>Infected Subjects</th>
<th>R5ef</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.24</td>
<td>Aggregated</td>
<td>8</td>
<td>0</td>
<td>Teunis</td>
</tr>
<tr>
<td>32.4</td>
<td>Aggregated</td>
<td>9</td>
<td>0</td>
<td>Teunis</td>
</tr>
<tr>
<td>324</td>
<td>Aggregated</td>
<td>9</td>
<td>3</td>
<td>Teunis</td>
</tr>
<tr>
<td>3240</td>
<td>Aggregated</td>
<td>3</td>
<td>2</td>
<td>Teunis</td>
</tr>
<tr>
<td>324,000</td>
<td>Aggregated</td>
<td>8</td>
<td>7</td>
<td>Teunis</td>
</tr>
<tr>
<td>3.24 * 10^6</td>
<td>Aggregated</td>
<td>7</td>
<td>3</td>
<td>Teunis</td>
</tr>
<tr>
<td>3.24 * 10^7</td>
<td>Aggregated</td>
<td>3</td>
<td>2</td>
<td>Teunis</td>
</tr>
<tr>
<td>3.24 * 10^8</td>
<td>Aggregated</td>
<td>6</td>
<td>5</td>
<td>Teunis</td>
</tr>
<tr>
<td>6.92 * 10^5</td>
<td>Disaggregated</td>
<td>8</td>
<td>3</td>
<td>Teunis</td>
</tr>
<tr>
<td>6.92 * 10^6</td>
<td>Disaggregated</td>
<td>18</td>
<td>14</td>
<td>Teunis</td>
</tr>
<tr>
<td>6.92 * 10^7</td>
<td>Disaggregated</td>
<td>1</td>
<td>1</td>
<td>Teunis</td>
</tr>
<tr>
<td>6.5 * 10^7</td>
<td>Disaggregated</td>
<td>13</td>
<td>10</td>
<td>Seitz</td>
</tr>
<tr>
<td>192</td>
<td>Aggregated</td>
<td>13</td>
<td>1</td>
<td>Atmar</td>
</tr>
<tr>
<td>1920</td>
<td>Aggregated</td>
<td>13</td>
<td>7</td>
<td>Atmar</td>
</tr>
<tr>
<td>19,200</td>
<td>Aggregated</td>
<td>8</td>
<td>7</td>
<td>Atmar</td>
</tr>
<tr>
<td>1.92 * 10^6</td>
<td>Aggregated</td>
<td>7</td>
<td>6</td>
<td>Atmar</td>
</tr>
<tr>
<td>2 * 10^7</td>
<td>Aggregated</td>
<td>23</td>
<td>16</td>
<td>Frenck</td>
</tr>
</tbody>
</table>
**Estimating $\alpha$ and $\beta$**

- Bayesian, using noninformative priors
- Account for aggregation in some inocula
- To avoid numerical issues, use confluent hypergeometric functions with R-Code provided by Dr. Peter Teunis
- Markov Chain Monte Carlo $\rightarrow$ sample of parameters $\alpha$, $\beta$, and $\mu$ (aggregation parm)
Where we came in!

• Try to reproduce Peter Teunis’ 2008 estimate.
  – Based on 11 dose groups (1\textsuperscript{st} 11 rows of our table)

• Not easy!
  – Some data in paper had 10x error.
  – Confluent hypergeometric function took some effort to check (R and MathCAD)

• Success! – Got same max likelihood parameter estimates.
  \begin{align*}
  \alpha &= 0.040 \\
  \beta &= 0.055 \\
  \text{mean aggregate size} &= 396.
  \end{align*}
But

• Likelihood contours and MCMC sample had “issues”
  – MCMC sample of size 10K had about 1.3K unique values (suggests poor mixing)
  – MCMC sample thinned where it shouldn’t (influence of normal prior)
  – Parameters $\alpha$ and $\beta$ were small. (Probability mass is concentrated near 0 and 1.)
The MCMC Sample
The Max Likelihood Beta Density

Probability Density

Probability of Infection, given exact dose is 1
It gets even more extreme.
So...the beta-Poisson doesn’t work well for Norovirus.

• MCMC sample thinned-out, but shouldn’t have with a truly non-informative prior.
• Bathtub is extremely deep.
  – Only 27% of the mass is shown in the figure!
  – 73% of the probability mass falls below 0.001 or above 0.999.
  – More than 1/3 of the probability mass falls below 0.000001 (1/million).
  – Another large fraction falls above 0.999999.

→ Most subjects are almost perfectly susceptible or almost perfectly immune to infection.
→ Why not exclude everything BUT 0 and 1? (Subjects would be either perfectly susceptible or perfectly immune.)
Fractional Poisson

• If we exclude everything but 0 and 1:
  • Each Se+ subject is either perfectly susceptible or perfectly immune.
  • Let $P = \text{fraction perfectly susceptible (parameter is 0)}$
  • $1 - P = \text{fraction perfectly immune (parameter is 1)}$
  • Perfectly susceptible subjects are infected if and only if they ingest at least one norovirus or norovirus aggregate.
Fractional Poisson is simple!

• Fraction P perfectly susceptible
• Poisson dose (with mean $\lambda$)
  – If large, probability of ingesting zero is nil.
  – If not large, probability of ingesting zero is Poisson probability of zero: $e^{-\lambda}$
  – Probability of ingesting one or more = $1 - e^{-\lambda}$
• Aggregation
  – Probability of ingesting zero = $e^{-\lambda/\mu}$
  – Probability of ingesting one or more = $1 - e^{-\lambda/\mu}$
  – Need to estimate mean aggregate size ($\mu$).
  – Aggregate size distribution is not important.
Estimating P and μ was simple!

- We found max likelihood solution:
  \[ P = 0.722 \text{ and mean aggregate size is 987} \]
- Likelihood is almost the same as for the beta-Poisson.
- AIC* favors fractional Poisson over beta-Poisson
- Likelihood contours and MCMC sample agree.
- Error structure is nearly normal.
- Estimation error is small (compared to beta-Poisson)

* AIC = Akaike Information Criterion
Code for Infection Probability (disaggregated dose)

**Beta-Poisson**

```r
дрbp <- function(alpha, beta, dose) 1 - (1 + dose/beta)^(-alpha)
```

**Fractional Poisson**

```r
др1f1 <- function(alpha, beta, dose) {
  if(dose<1E-4) return (alpha*dose/(alpha+beta))  # Corrected
  if(alpha>1E3 && beta < alpha/100) return (1-exp(-dose))
  if(alpha>1E2 && beta>1E5) return (1-exp(-dose*alpha/beta))
  if(alpha>1E1 && beta>1E5 && dose*alpha/beta>10) return (дрbp(alpha, beta, dose))
  if(alpha>1 && beta>alpha*20 && dose>10) return (дрbp(alpha, beta, dose))
  if(alpha>1 && beta>alpha && dose>50) return (дрbp(alpha, beta, dose))
  if(alpha<1 && beta>alpha*50) return (дрbp(alpha, beta, dose))
  if(alpha<0.1 && beta>alpha*20) return (дрbp(alpha, beta, dose))
  if(abs(round(alpha)-alpha)<1E-4) alpha=1.0001*alpha
  if(abs(round(beta)-beta)<1E-4) beta=1.0001*beta
  return (1-hyperg_1F1(alpha, alpha+beta, -dose))
}
```
Normal contours.
Well-behaved MCMC sample.
Estimated Infection Probability

Beta-Poisson

Fractional Poisson
95% prediction intervals at Dose == 1

- Beta-Poisson: 0.019 to 0.76
- Fractional Poisson: 0.63 to 0.8

But: We haven’t ruled out the beta model. We really have no low dose data, so we wouldn’t be surprised if new low dose data were to favor the beta model.
Imagine a new study...

What if 10 Se+ individuals each ingested exactly 1 virion and none was infected? That would not be likely under the fractional model. Likelihood would favor the beta.

If instead, 5 to 10 were infected, that would be consistent with both models. Again, the fractional Poisson would be the model of choice, due to having fewer parameters.
Other pathogens, for which the fractional model may work:

- Cryptosporidium
- Campylobacter jejuni
- Shigella flexneri
- Rotavirus

Data from these studies show dose-response functions that may have peaked and/or MCMC samples that include cases where beta distribution parameters can be arbitrarily small.
Acknowledgements

• Thanks to Peter Teunis for
  – Providing his MCMC sample
  – Providing R-code for computing beta-Poisson probabilities

• Thanks to Sharon Nappier and Phil Berger for pulling me into this and making it fun (e.g., slide 3’s graphics).
Disclaimer

Views expressed in this presentation are the author’s and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.