

## Bayesian models of longitudinal IFN resistance

To create a simple model of the longitudinal dynamics of IFN $\alpha$ 2 and IFN $\beta$  IC<sub>50</sub> were each modeled using a Bayesian change point hierarchical model. The model is based on a segmented regression of the log IC<sub>50</sub> where:

- each participant has an acute level of resistance at initial infection drawn from separate population distributions for typical, non- or fast progressors
- each participant has a drop (or rise) in resistance from acute levels drawn from separate population distributions for typical, non- or fast progressors
- each participant has a time to nadir (or zenith) drawn from a shared population distribution
- resistance changes linearly from onset of symptoms to time of nadir
- after nadir interferon resistance depends linearly on change in CD4 count from the level found at nadir with the effect of CD4 for each participant drawn from separate population distributions for typical, non- or fast progressors
- CD4 count was linearly interpolated between observations

The log IC<sub>50</sub> observation from each viral isolate  $i$  was modeled as a normal distribution  $IC50_i \sim \text{Normal}(\mu_i, \sigma)$  with mean  $\mu_i$  where:

$$\mu_i = \begin{cases} \alpha_{\text{person}_i} + \delta_{\text{person}_i} \times \frac{\text{time}_i}{s_{\text{person}_i}} & \text{if } \text{time}_i < s_{\text{person}_i} \\ \alpha_{\text{person}_i} + \delta_{\text{person}_i} + \beta_{\text{person}_i} \times (\text{CD4}_{\text{person}_i, \text{time}_i} - \text{CD4}_{\text{person}_i, s_{\text{person}_i}}) & \text{if } \text{time}_i \geq s_{\text{person}_i} \end{cases}$$

where the parameters  $\alpha_j$  represent the level of interferon resistance at symptom onset,  $\delta_j$  represents the change from symptom onset to nadir and  $s_j$  represents the time of nadir in patient  $j$ . Patient data is represented by  $\text{time}_i$  and  $\text{patient}_i$  corresponding to respectively the time and participant from which isolate  $i$  was collected and  $\text{CD4}_{j,k}$  containing the estimated CD4 count for patient  $j$  at time  $k$  and  $\text{progression}_j$  is the progression type (fast/non/typical)

for participant  $j$ . The hierarchical probabilities for these parameters were:

$$\begin{aligned}
\sigma &\sim \text{Gamma}(1, 0.1) \\
\alpha_j &\sim \begin{cases} \text{Normal}(\theta_{\alpha, \text{typical}}, \tau_{\alpha}) & \text{if progression}_j = \text{typical} \\ \text{Normal}(\theta_{\alpha, \text{typical}} + \theta_{\alpha, \text{fast}}, \tau_{\alpha}) & \text{if progression}_j = \text{fast} \\ \text{Normal}(\theta_{\alpha, \text{typical}} + \theta_{\alpha, \text{non}}, \tau_{\alpha}) & \text{if progression}_j = \text{non} \end{cases} \\
\delta_j &\sim \begin{cases} \text{Normal}(\theta_{\delta, \text{typical}}, \tau_{\delta}) & \text{if progression}_j = \text{typical} \\ \text{Normal}(\theta_{\delta, \text{typical}} + \theta_{\delta, \text{fast}}, \tau_{\delta}) & \text{if progression}_j = \text{fast} \\ \text{Normal}(\theta_{\delta, \text{typical}} + \theta_{\delta, \text{non}}, \tau_{\delta}) & \text{if progression}_j = \text{non} \end{cases} \\
s_j &\sim \text{NegativeBinomial}(\theta_s, \tau_s) \\
\beta_j &\sim \begin{cases} \text{Normal}(\theta_{\beta, \text{typical}}, \tau_{\beta}) & \text{if progression}_j \in \text{typical, non} \\ \text{Normal}(\theta_{\beta, \text{typical}} + \theta_{\beta, \text{fast}}, \tau_{\beta}) & \text{if progression}_j = \text{fast} \end{cases}
\end{aligned}$$

where  $j$  indicates each participant and  $\text{NegativeBinomial}(x, y)$  represents a negative binomial distribution parameterized such that the expected value is  $x$  and the variance is  $x + \frac{x^2}{y}$ . All hyperparameters were given prior probabilities of  $\theta_x \sim \text{Normal}(0, 10)$  for parameters representing the means of a distribution and  $\tau_x \sim \text{Gamma}(1, 0.1)$  for parameters representing standard deviations other than  $\theta_{\alpha, \text{typical}}$  and  $\theta_s$  which were given a flat prior and  $\tau_s \sim \text{Cauchy}(0, 10)$ .

For computational efficiency, the nadir time parameter  $s$  was discretized to weekly intervals, assumed to fall within 1 to 150 weeks after symptom onset and marginalized out of the joint probability:

$$p(\text{IC50}, \dots) = p(\dots) \prod_{i=1}^n \sum_{k=1}^{150} \text{Normal}(\text{IC50}_i | \mu_{i,s}, \sigma) \text{NegativeBinomial}(s | \theta_s, \tau_s)$$

where  $\dots$  represents all parameters other than  $s$  and  $\mu_{i,s}$  is defined the same as  $\mu_i$ :

$$\mu_{i,s} = \begin{cases} \alpha_{\text{person}_i} + \frac{\text{time}_i}{s} \delta_{\text{person}_i} & \text{if time}_i < s \\ \alpha_{\text{person}_i} + \delta_{\text{person}_i} + \beta_{\text{person}_i} (\text{CD4}_{\text{person}_i, \text{time}_i} - \text{CD4}_{\text{person}_i, s}) & \text{if time}_i \geq s \end{cases}$$

## Bayesian models of outgrowth and rebound resistance

To compare virus isolated during rebound, ART-suppression (QVOA), acute and chronic stages of infection, difference between the  $\text{IFN}\alpha_2$  and  $\text{IFN}\beta$   $\text{IC}_{50}$  were modeled using a Bayesian hierarchical model. The model is based on the assumptions that:

- Isolates within a patient were assumed to share some similarity in IFN resistance.
- The mean  $\text{IC}_{50}$  level within each person for a given virus type was assumed to be drawn from a shared population distribution for that type.

- QVOA virus were separated into to two populations; “pre” a group composed of QVOA viruses isolated from patients without treatment interruptions (ATI) or prior to ATI and “post” QVOA viruses isolated from patients following ATI.
- In both QVOA populations, the viruses may to be composed of a population found in QVOAs plus an additional admixture of virus similar to those observed during rebound i.e. potential reseeding after ATI. We model a mixture in both pre- and post-treatment so that we can asses differences in inferred mixture proportion between the two populations.

The log IC<sub>50</sub> observation from each viral isolate  $i$  from acute, chronic and rebound isolates was modeled as a normal distribution:

$$\text{IC50}_i \sim \text{Normal}(\mu_{\text{type}_i, \text{person}_i}, \sigma_{\text{type}_j})$$

with the mean resistance for isolate type  $j$  from person  $k$ :

$$\mu_{j,k} \sim \begin{cases} \text{Normal}(\alpha_k + \beta_{\text{batch}} \text{batch}_k, \psi_j) & \text{if } j = \text{acute} \\ \text{Normal}(\alpha_k + \beta_{j,k} + \beta_{\text{batch}} \text{batch}_k + \beta_{\text{IFN}} \text{IFN}_k, \psi_j) & \text{if } j = \text{rebound} \\ \text{Normal}(\alpha_k + \beta_{j,k} + \beta_{\text{batch}} \text{batch}_k, \psi_j) & \text{otherwise} \end{cases}$$

where  $\text{type}_i$  indicates whether isolate  $i$  was isolated during acute, chronic or rebound infection from participant  $\text{person}_i$ ,  $\text{batch}_k$  indicates when isolates from person  $k$  were tested in Iyer et al. 2017 (to account for variation in IFN batches from this study) and  $\text{IFN}_k$  indicates when person  $k$  was given exogenous IFN $\alpha$ 2 during treatment interruption. Parameters are included for the mean resistance level during acute infection for each person  $\alpha_k$ , standard deviation of isolates of type  $j$  within a person  $\sigma_j$ , standard deviation of mean resistance for type  $j$  isolates among people  $\psi_j$ , change from acute levels in isolates of type  $j$  in a given patient  $\beta_{j,k}$ , the effects of exogenous IFN treatment  $\beta_{\text{IFN}}$  and batch to batch variation in IFN in isolates previously assayed by Iyer et al. 2017  $\beta_{\text{batch}}$ .

For QVOA isolates, the IC<sub>50</sub> was modeled as a mixture of two populations such that:

$$\begin{aligned} p(\text{IC50}_i | \mu_{\text{QVOA}, \text{person}_i}, \sigma_{\text{QVOA}}, \mu_{\text{rebound}, \text{person}_i}, \sigma_{\text{rebound}}, \phi_{\text{prePost}_i}) = \\ \phi_{\text{prePost}_i} \text{Normal}(\text{IC50}_i | \mu_{\text{rebound}, \text{person}_i}, \sigma_{\text{rebound}}) \\ + (1 - \phi_{\text{prePost}_i}) \text{Normal}(\text{IC50}_i | \mu_{\text{QVOA}, \text{person}_i}, \sigma_{\text{QVOA}}) \end{aligned}$$

where  $\text{prePost}_i$  indicates whether isolate  $i$  was isolated pre- or post-ATI and  $\phi_{\text{pre}}$  and  $\phi_{\text{post}}$  represent the proportion of rebound-like virus present in pre- and post-ATI QVOA isolates.

The hierarchical parameter priors were modeled as:

$$\begin{aligned} \sigma_j &\sim \text{Gamma}(1, 0.1) \\ \psi_j &\sim \text{Gamma}(1, 0.1) \\ \phi_{\text{pre}} &\sim \text{Uniform}(0, 1) \\ \phi_{\text{post}} &\sim \text{Uniform}(0, 1) \end{aligned}$$

where  $j$  indicates the isolate type (acute, chronic, qvoa, rebound).  $\alpha_k$ ,  $\beta_{j,k}$ ,  $\beta_{\text{IFN}}$  and  $\beta_{\text{batch}}$  were all given flat priors.