Supplementary Information

## Appendix 1 - Creating A Symbolic Link

1. Download an EpiFusion jar file and place it somewhere safe on your system, for example /usr/local/bin.
2. Open a terminal window and type

* echo $PATH

1. Navigate to one of the filepaths that appears.
2. Create a bash script that will run EpiFusion and pass arguments to it. You can do this with a text editor like ‘nano’ or ‘vim’. The bash script can be called anything, but it should contain the following content:

* #!/bin/bash  
  java -jar /absolute/path/to/EpiFusion/jar/file.jar "$@"

1. Create a symbolic link between this script and the keyword ‘EpiFusion’. You can do that by typing the following command:

* ln -s /path/to/wherever/the/bash/script/was/made EpiFusion

Now you should be able to call EpiFusion from anywhere in your system.

## Appendix 2 - Full EpiFusion XML Breakdown

### Loggers

The loggers section dictates the program output,

#### fileBase

The name of the folder to be created for for output files (<fileBase>). Unless you specify a filepath in fileBase, the output files will be written to your working directory. This can be anything, but avoid the usual folder naming trouble makers (spaces, slashes, etc).

#### logEvery

This sets the frequency at which the program logs the state of the MCMC to the output files (and prints to console).

### Data

The data section is for providing case incidence data, a phylogenetic tree, and how you wish to weight their contribution to the model.

#### Incidence

You can provide incidence to the model inside the <incidence> tag.

<incidence>  
 <incidenceVals>1 3 1 5 18 35 40 25 24 13 2 0 1 0</incidenceVals>  
 <incidenceTimes type = "every">7<incidenceTimes>  
</incidence>

##### IncidenceVals

The values go in the <incidenceVals> tag, ordered forwards in time from the beginning of the modelled time period

##### IncidenceTimes

The incidenceTimes tag can be used to specify the times of incidence observations forwards in time from the beginning of the model time period. There are a few options for specifying the times with the type attribute inside the tag - either specifying the interval between observations or specifying exact times that incidence was recorded:

<incidenceTimes type = "every">7<incidenceTimes>

For example above we specify that incidence is recorded every 7 days, starting on day 7.

<incidenceTimes type = "exact">10 14 25 31 56<incidenceTimes>

And here we provide the exact times that incidence data was recorded. Note that if you use this option, the length of incidenceTimes should match incidenceValues.

#### Tree

Provide a timescaled phylogenetic tree to the model using the <tree> tag. Currently, it is necessary to have node a leaf labels formatted ending with square brackets with the node times in these brackets (e.g. node\_1[0.3442]). This requires some preprocessing, the ape in R package does the trick, or you can use the prep\_epifusion\_data() function in EpiFusionUtilities that can do this conversion for you too. We recognise this is inconvenient and removing this requirement to allow normal nexus trees is on the top of our TODO list.

<tree>((((((((leaf\_407[70.10236230370634]:19.09030831,leaf\_377[71.67903395633373]:20.66697996)node\_23[51.0120539946585]:16.72849832,(leaf\_530[64.29019242068087]:28.60128838,(((leaf\_678[55.836692533807465]:12.44871907,leaf\_692[54.95812431493469]:11.57015085)node\_52[43.38797346813441]:0.385599307,leaf\_640[58.57019798061809]:15.56782382)node\_51[43.00237416117269]:4.149624169,leaf\_223[80.20988528843057]:41.3571353)node\_49[38.852749991991104]:3.163845956)node\_32[35.68890403623855]:1.405348363)node\_9[34.28355567352641]:5.723132688,((leaf\_242[78.62221627030786]:37.6032997,((leaf\_153[85.38189406711462]:30.4653199,leaf\_75[95.31697617294901]:40.400402)node\_112[54.91657417003654]:9.036630564,leaf\_111[90.02540774077697]:44.14546413)node\_108[45.879943605975534]:4.861027036)node\_87[41.018916570389365]:6.268352098,((leaf\_28[110.31768492437661]:56.8016204,leaf\_445[68.36948591902012]:14.8534214)node\_161[53.51606452349725]:6.970120811,leaf\_752[48.45392169632879]:1.907977984)node\_160[46.54594371207695]:11.79537924)node\_84[34.750564472466806]:6.190141487)node\_8[28.560422985624168]:4.876221701,(((((leaf\_74[95.8576156329268]:37.00079705,leaf\_396[70.55955806481323]:11.70273948)node\_192[58.85681858553985]:10.32855116,leaf\_662[57.072280245746875]:8.544012822)node\_190[48.52826742337909]:5.282985936,leaf\_109[90.08763067992984]:46.84234919)node\_187[43.24528148741703]:6.403050181,leaf\_102[90.85197148216028]:54.00974018)node\_185[36.84223130663702]:1.46060558,((leaf\_70[96.23083229433816]:57.38098621,leaf\_596[60.77116595386879]:21.92131987)node\_223[38.84984607971327]:0.1026768434,(leaf\_159[84.84128196112528]:30.74399116,leaf\_166[84.10666840469251]:30.0093776)node\_240[54.0972908003288]:15.35012156)node\_222[38.74716923627249]:3.36554351)node\_176[35.38162572640305]:11.69742444)node\_7[23.68420128510857]:11.46063414,(((leaf\_160[84.82844619112666]:60.75968714,((((leaf\_64[98.73242319027953]:58.21388631,leaf\_681[55.58081691177014]:15.06228003)node\_294[40.51853687799632]:1.735152486,leaf\_590[61.148684429556546]:22.36530004)node\_293[38.783384391686475]:0.551139453,leaf\_98[91.67047000399187]:53.43822507)node\_292[38.23224493867514]:10.97801371,leaf\_791[43.377517040053526]:16.12328581)node\_289[27.25423123067509]:3.185472183)node\_248[24.068759047729554]:0.4997283961,(leaf\_553[63.34492352254521]:28.29037819,leaf\_99[91.60953189160091]:56.55498656)node\_329[35.05454533491508]:11.48551468)node\_247[23.56903065162453]:7.719268377,((leaf\_309[74.58310015128357]:13.36541254,leaf\_55[101.81985847090435]:40.60217086)node\_362[61.21768760926006]:34.52413937,leaf\_424[69.34790109229266]:42.65435285)node\_359[26.69354824266421]:10.84378597)node\_246[15.849762274390272]:3.62619513)node\_5[12.223567144247692]:3.195696408,(((leaf\_162[84.40937076572243]:31.50911351,leaf\_324[73.97434087205923]:21.07408361)node\_395[52.90025725951645]:9.663695667,leaf\_507[65.4925795095511]:22.25601792)node\_388[43.23656159282253]:12.85871416,(((leaf\_667[56.73582187536703]:2.027004712,leaf\_271[76.9847635381094]:22.27594637)node\_442[54.70881716344273]:1.414390225,leaf\_366[72.0731573444004]:18.77873041)node\_441[53.294426938013345]:12.38434243,leaf\_292[75.68955369039529]:34.77946918)node\_431[40.91008450554362]:10.53223707)node\_385[30.377847436622286]:21.3499767)node\_4[9.027870735773057]:2.202856024,((((((((leaf\_59[100.08067285788199]:50.16113194,(leaf\_122[88.25148926376218]:36.51609694,leaf\_481[66.5178885482595]:14.78249623)node\_490[51.73539232236806]:1.815851402)node\_488[49.91954092033317]:5.658838175,leaf\_783[44.504551838421605]:0.2438490932)node\_485[44.260702745210416]:3.834595734,leaf\_718[52.73400913067007]:12.30790212)node\_482[40.42610701140656]:0.3727987892,leaf\_137[86.87820106224541]:46.82489284)node\_481[40.053308222256206]:1.541827426,leaf\_679[55.73806406067391]:17.22658326)node\_480[38.51148079662553]:11.31956052,((leaf\_415[69.74802935404291]:24.16809076,leaf\_542[63.98421692880167]:18.40427833)node\_534[45.579938595143936]:16.55024436,leaf\_494[65.88016294241126]:36.8504687)node\_525[29.02969423779227]:1.837773964)node\_469[27.19192027355947]:4.784286462,(((((leaf\_240[78.7251493051829]:36.09533547,leaf\_733[51.27464829974004]:8.644834469)node\_572[42.62981383087519]:3.351534295,leaf\_565[62.54393520248931]:23.26565567)node\_571[39.27827953617371]:10.94774507,((((leaf\_544[63.77673952279264]:26.31061793,leaf\_363[72.15967419706526]:34.69355261)node\_611[37.46612159103572]:0.04411065085,leaf\_41[105.11850117358156]:67.69649023)node\_610[37.4220109401828]:2.236590466,((leaf\_505[65.51898007568552]:14.55632746,leaf\_452[67.98768873895207]:17.02503612)node\_662[50.96265262067672]:1.536519265,leaf\_466[67.2387130185725]:17.81257966)node\_661[49.42613335594053]:14.24071288)node\_609[35.18542047400971]:4.220130407,leaf\_774[45.24321718225812]:14.27792712)node\_606[30.965290067180977]:2.634755598)node\_570[28.330534468731727]:0.9269283664,(((((leaf\_140[86.67125153426835]:46.3559461,leaf\_524[64.58552017616272]:24.27021474)node\_700[40.315305433596436]:2.215938869,((leaf\_334[73.56013467469849]:10.73098258,leaf\_517[65.14498837179052]:2.315836279)node\_731[62.829152092327995]:15.4593817,leaf\_307[74.69334748527966]:27.32357709)node\_728[47.36977039079453]:9.270403827)node\_698[38.099366564151296]:3.31115576,leaf\_444[68.42688854802711]:33.63867774)node\_697[34.78821080397044]:2.208122614,(leaf\_564[62.62160623104026]:17.51198355,leaf\_763[47.149467892186756]:2.039845215)node\_744[45.109622677313816]:12.52953449)node\_696[32.580088189594434]:3.733445308,(((leaf\_63[99.03748147442674]:36.16276457,leaf\_183[82.93567799222579]:20.06096108)node\_760[62.87471690918196]:16.65804619,(leaf\_65[97.92430834876784]:41.86612862,leaf\_209[80.97839091724356]:24.92021119)node\_762[56.058179730263774]:9.84150901)node\_758[46.216670720407905]:13.3159056,leaf\_811[36.537443151630136]:3.63667803)node\_754[32.90076512176364]:4.05412224)node\_694[28.846642881868824]:1.44303678)node\_569[27.403606102349244]:1.228209889,((leaf\_319[74.20982901626131]:24.16887906,(leaf\_492[66.0362745797389]:12.06101312,(leaf\_451[68.04527961439327]:10.70473666,leaf\_487[66.15872756629061]:8.818184615)node\_793[57.34054295105745]:3.365281488)node\_790[53.975261462789035]:3.934311504)node\_776[50.04094995889265]:4.478634208,leaf\_94[92.54528570216107]:46.98296995)node\_775[45.562315751008036]:19.38691954)node\_568[26.175396213176917]:3.767762401)node\_466[22.40763381199003]:2.284852332,(leaf\_670[56.498190314744825]:6.555477923,leaf\_304[74.78607366094431]:24.84336127)node\_811[49.94271239223166]:29.81993091)node\_465[20.122781479584827]:13.29776677)node\_3[6.825014711946831]:0.3910339013,leaf\_212[80.7513012802764]:74.31732047)node\_2[6.433980810655798]:6.433980810655798;  
 </tree>

#### Weighting

The data block also allows you to set the ‘influence’ each set of data has over the inferred epidemic trajectories, using the epicontrib and changetimes parameters, however ***we currently do not recommend setting these parameters as anything but the defaults (0.5 and 0)***. These default values essentially make it so that each dataset is contributing equally, and this is all we have validated the model for so far.

<epicontrib>0.5</epicontrib>  
<changetimes>0</changetimes>

### Analysis

The analysis block is another block you probably won’t have to change from the default. It will usually look like this:

<analysis>  
 <type>looseformbeta</type>  
 <startTime>null</startTime>  
 <endTime>null</endTime>  
 </analysis>

Most analyses will use looseformbeta as the type. This means you don’t want to set any expectations for the movements beta (the daily rate of infection) will make. So, each particle in the particle filter will each have their own beta, which will vary over time in a random walk and yield a unique beta trajectory (you can parameterise the nature of this random walk in priors). You can learn more about the different options for fitting beta through time [here](https://github.com/ciarajudge/EpiFusion/wiki/Options-for-Fitting-Beta).

You can also specify the start and end time of the analysis (in days) if you have data covering a timeframe longer than what you are interested in - this is mainly for epi only analyses.

### Model

In this block you can customise the model a bit more - as we add improvements to EpiFusion we intend to make it flexible and modular to allow more user control over the model. For now, you can use this block to decide between epidemiological observation models. There are currently two options:

#### Poisson

<model>  
 <epiObservationModel>poisson</epiObservationModel>  
 </model>

#### Negative Binomial

The negative binomial can be used as a special case of the poisson to account for overdispersion in your data. Currently the overdispersion parameter must be set manually - eventually we hope to offer the ability to infer it.

<model>  
 <epiObservationModel>negbinom</epiObservationModel>  
 <overdispersion>10.0</overdispersion>  
 </model>

### Parameters

There are **a lot** of parameters to choose from for your analysis, and this and the priors block are the ones you will likely be editing the most - so get familiar. Here’s what it looks like:

<parameters>  
 <epiOnly>false</epiOnly>  
 <phyloOnly>false</phyloOnly>  
 <numParticles>100</numParticles>  
 <numSteps>2000</numSteps>  
 <numThreads>8</numThreads>  
 <numChains>4</numChains>  
 <stepCoefficient>0.05</stepCoefficient>  
 <resampleEvery>7</resampleEvery>  
 <segmentedDays>true</segmentedDays>  
 <maxEpidemicSize>50000</maxEpidemicSize>  
 <samplingsAsRemovals>1</samplingsAsRemovals>  
 <generationPMF>0.03227895 0.04600192 0.05888228 0.06604339 0.06944715 0.07072715 0.06937254 0.06716382 0.06340303 0.05980526 0.05536372 0.05122294 0.04693064 0.04298847 0.03875701 0.03488946 0.03107127 0.02825642 0.02513621 0.02228003 0.01997833 </generationPMF>  
 </parameters>

Now let’s go through each component in turn.

#### epiOnly

Set it to true if you want to run an case incidence only analysis. If it, and phyloOnly are both false, a combined analysis will be run.

#### phyloOnly

Set it to true if you want to run a tree only analysis. If it, and epiOnly are both false, a combined analysis will be run.

#### numParticles

Does what it says on the tin: numParticles sets the number of particles for the particle filter. 100-200 particles usually works fine, if you are expecting unpredictable movements it’s usually helpful to aim towards the higher end.

#### numSteps

Number of steps of the MCMC. The particle filter does a lot of heavy lifting fitting wise, so 2000-5000 steps often does the trick.

#### numThreads

Number of threads to use during the parallelised particle filter steps.

#### numChains

Number of MCMC chains to run. Standard practice is around 4.

#### stepCoefficient

Coefficient of the MCMC step sizes for generating new proposals. Currently the MCMC sampler is basic Metropolis Hastings so you might have to mess around with this step-size to find what works best.

#### resampleEvery

This sets the number of days between each particle resampling. 7 days has worked nicely for us so far, and gels nicely with weekly case incidence counts.

#### segmentedDays

This makes a minor change to the phylogenetic likelihood implementation. When it is false, the phylogenetic weighting is done in discrete daily increments, but if you have a tree with a lot happening (multiple internal or external nodes per day) then <segmentedDays>true</segmentedDays> breaks down the days into slightly finer chunks which can help with your MCMC acceptance rate when prevalence is low. It’s generally better safe than sorry to keep segmentedDays as true, it just marginally slows down the analysis (very slightly).

#### maxEpidemicSize

This is way to limit the size of the modelled epidemic if required.

#### samplingsAsRemovals

If you are modelling sampling events on the tree as removals from the infected population, this should be 1. Otherwise set it to 0.

### Priors

The priors block is where you provide priors and other parameterisations for the rates of the particle filter process model. You can provide a number of different priors, but you definitely must include a gamma, psi and phi. All priors need two things: whether or not the rate has a step-change (stepchange), and then details on how the rate is distributed (disttype and associated info). For example, here is a gamma parameterisation where the rate stays constant over time (no step-change) and a truncated normal distribution with mean 0.143, standard deviation 0.03 and lower bound of 0.0.

<gamma>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.143</mean>  
 <standarddev>0.03</standarddev>  
 <lowerbound>0.0</lowerbound>  
 <stepchange>false</stepchange>  
 </gamma>

#### Distribution Options

There are a number of different distribution options available with more being added all the time. Here is what’s currently available: 1. Normal 2. Truncated Normal 3. Poisson 4. Uniform 5. Uniform Discrete 6. Fixed Parameter (this is if you want to fix a parameter, i.e. not infer it)

And here’s an example of how each of them could be parameterised:

<exampleparam>  
 <disttype>Normal</disttype>  
 <mean>0.143</mean>  
 <standarddev>0.03</standarddev>  
 <stepchange>false</stepchange>  
 </exampleparam>  
  
 <exampleparam>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.143</mean>  
 <standarddev>0.03</standarddev>  
 <lowerbound>0.0</lowerbound>  
 <stepchange>false</stepchange>  
 </exampleparam>  
  
 <exampleparam>  
 <disttype>Poisson</disttype>  
 <mean>500</mean>  
 <stepchange>false</stepchange>  
 </exampleparam>  
  
 <exampleparam>  
 <disttype>Uniform</disttype>  
 <min>0.1</min>  
 <max>0.8</max>  
 <stepchange>false</stepchange>  
 </exampleparam>  
  
 <exampleparam>  
 <disttype>UniformDiscrete</disttype>  
 <min>400</min>  
 <max>600</max>  
 <stepchange>false</stepchange>  
 </exampleparam>  
  
 <exampleparam>  
 <disttype>FixedParameter</disttype>  
 <value>0.15</value>  
 <stepchange>false</stepchange>  
 </exampleparam>

#### Parameterising beta

For looseformbeta analyses types, which we recommend using most of the time, you’ll also need to provide an initialBeta parameter and betaJitter, which gives particles their initial beta value, and specifies the ‘freedom’ of the random walk. Those will look like this:

<initialBeta>  
 <stepchange>false</stepchange>  
 <disttype>Uniform</disttype>  
 <min>0.05</min>  
 <max>0.3</max>  
 </initialBeta>  
 <betaJitter>  
 <stepchange>false</stepchange>  
 <disttype>Uniform</disttype>  
 <min>0.001</min>  
 <max>0.05</max>  
 </betaJitter>

If you are fitting beta as an inverse logistic function, it is parameterised a little differently (you provide priors for the three parameters of the inverse logistic curve:

<a>  
 <stepchange>false</stepchange>  
 <disttype>Normal</disttype>  
 <mean>0.047</mean>  
 <standarddev>0.01</standarddev>  
 </a>  
 <b>  
 <stepchange>false</stepchange>  
 <disttype>Normal</disttype>  
 <mean>-0.06</mean>  
 <standarddev>0.01</standarddev>  
 </b>  
 <c>  
 <stepchange>false</stepchange>  
 <disttype>Normal</disttype>  
 <mean>0.4</mean>  
 <standarddev>0.1</standarddev>  
 </c>

#### Introducing rate changes

The previous examples showed had the rates as constant over time (stepchange = false). When you introduce a step-change by changing the parameter to true, you then need to set priors for each interval, and the interval change times. This moves disttype and the associated tags inside new XML nodes called changetime and distribs. The best way to explain this is with an example. First lets look at this parameterisation of psi, where psi is constant over time and has a truncated normal distribution:

<psi>  
 <stepchange>false</stepchange>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.00025</mean>  
 <standarddev>0.00005</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </psi>

Let’s introduce a step-change where we expect psi to increase 10x at a specific time (day 35). First we set stepchange to true, then add changetime and distribs tags, inside which we set individual priors for the rates in each interval, and the interval times. These are now wrapped in tags <x[n]>, starting at n=0 and counting up for every extra interval or rate you are adding. You can specify as many changetimes as you like, but you should have one more distribs element than the number of changetime elements.

<psi>  
 <stepchange>true</stepchange>  
 <changetime>  
 <x0>  
 <disttype>FixedParameter</disttype>  
 <value>35</value>  
 </x0>  
 </changetime>  
 <distribs>  
 <x0>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.00025</mean>  
 <standarddev>0.00005</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </x0>  
 <x1>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.0025</mean>  
 <standarddev>0.0005</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </x1>  
 </distribs>  
 </psi>

Finally, let’s assume we don’t know the time of the step-change, and we want to infer it. The result is the exact same as above, but changetime <x0> is now no longer a FixedParameter, but instead a Poisson.

<psi>  
 <stepchange>true</stepchange>  
 <changetime>  
 <x0>  
 <disttype>Poisson</disttype>  
 <mean>35</mean>  
 </x0>  
 </changetime>  
 <distribs>  
 <x0>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.00025</mean>  
 <standarddev>0.00005</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </x0>  
 <x1>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.0025</mean>  
 <standarddev>0.0005</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </x1>  
 </distribs>  
 </psi>

## Appendix 3 - Main EpiFusionUtilities Functions

### load\_raw\_epifusion("path\_to\_outputfolder")

* *Description:* Loads an R list with the full posterior trajectory samples, parameter samples, likelihoods and acceptance rates for all chains.
* *Input:* A filepath to an EpiFusion outputs folder
* *Output:* A large list with the full EpiFusion output in an R compatible format.

raw\_epifusion\_output <- load\_raw\_epifusion('epifusionoutputs/')

### plot\_likelihood\_trace(raw\_epifusion\_output)

* *Description:* Plots the likelihood trace for all chains in an EpiFusion analysis
* *Input:* Raw epifusion output object generated by load\_raw\_epifusion()
* *Output:* A plot of the likelihood trace coloured by chain.

plot\_likelihood\_trace(raw\_epifusion\_output)

### extract\_posterior\_epifusion(raw\_epifusion\_output, burn\_in)

* *Description:* Extracts the final epifusion posterior samples, discarding a certain proportion at the beginning of each chain as burn-in
* *Input:* Raw epifusion output object generated by load\_raw\_epifusion() and proportion of each chain to discard as burn-in
* *Output:* A large list with the EpiFusion posterior samples aggregated across chains (with burn-in discarded). Also has trajectory means and HPDs, and parameter convergence statistics.

extract\_posterior\_epifusion(raw\_epifusion\_output, 0.1)

## Appendix 4 - Paired Sampling Infrastructure

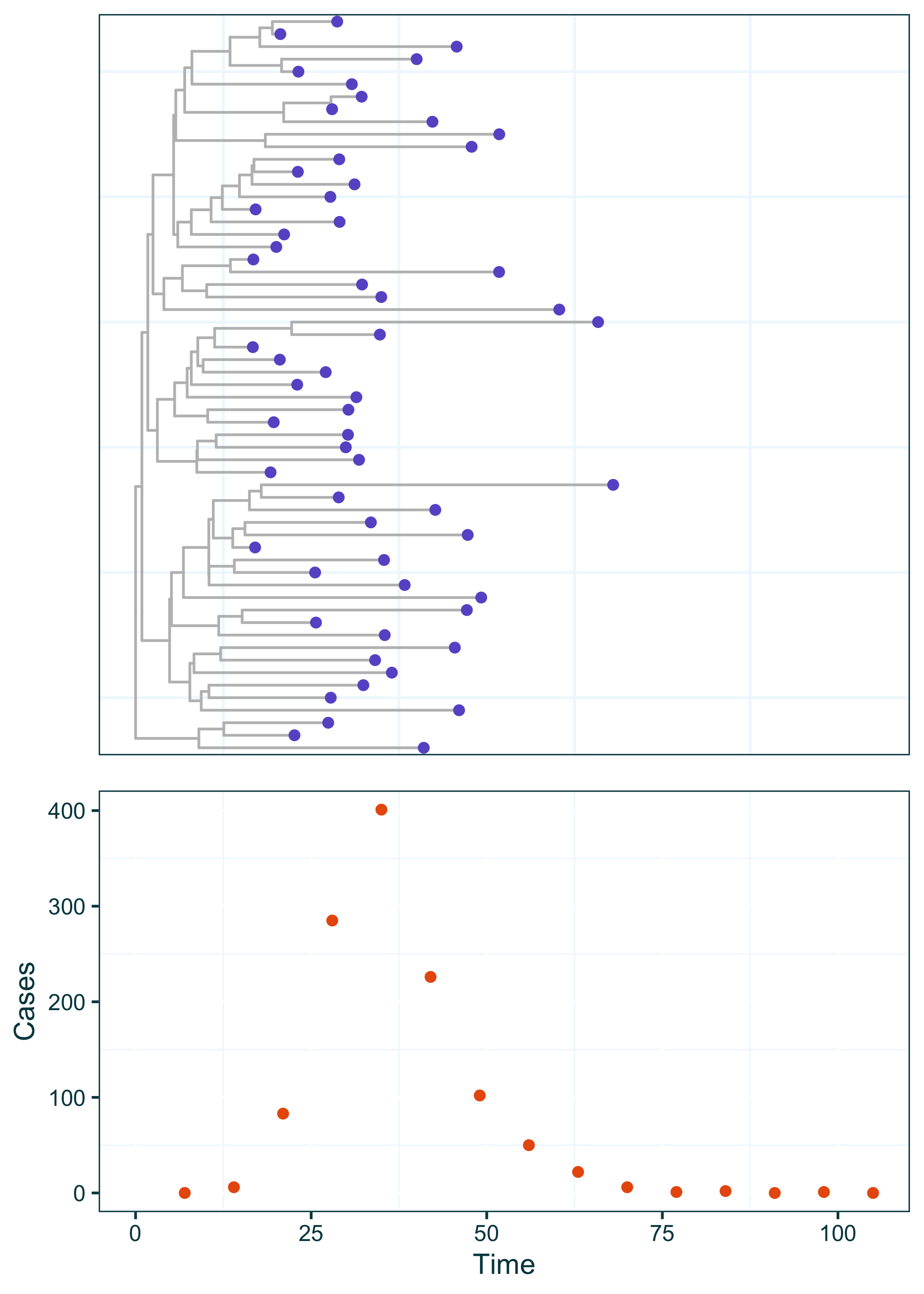
Specifying the case and genomic sequence sampling rates for EpiFusion is an important aspect of the parameterisation, however it should be noted that the these should be somewhat linked i.e. their proportion of the total infections will be related to their proportion of each other. With this in mind it is possible to parameterise the psi (genomic sequencing proportion of all cases) parameter as being linked to the phi parameterisation, where the fraction of sampled sequences as a proportion of cases over time is used in conjunction with the phi parameterisation, in place of specifying these parameters separately.

Consider an analysis with the following data chunk in it’s XML file:

<data>  
 <incidence>  
 <incidenceVals>0 6 83 285 401 226 102 50 22 6 1 2 0 1 0</incidenceVals>  
 <incidenceTimes type="every">7</incidenceTimes>  
 </incidence>  
 <tree>(((leaf\_978[24.289092735613732]:10.05812796,leaf\_751[29.09361836211882]:14.86265359)node\_14[14.230964773732369]:3.553266042,leaf\_172[42.69772739600777]:32.02002866)node\_4[10.67769873203644]:9.009220774,(((((leaf\_727[29.456732285001838]:17.34678626,leaf\_444[34.098590228053204]:21.9886442)node\_160[12.109946023683246]:1.093924779,leaf\_98[47.72409501869759]:36.70807377)node\_159[11.016021244423117]:1.619071921,(leaf\_292[38.15303138198247]:28.16886388,(leaf\_379[35.76908018744906]:21.99507627.........node\_968[7.674601453741431]:0.5900236866)node\_900[7.084577767125148]:2.936346776,(leaf\_11[61.99365320703872]:56.29610808,((leaf\_348[36.65407632226596]:24.86354942,leaf\_457[33.91669350511818]:22.1261666)node\_1100[11.790526905301487]:3.459163323,(leaf\_46[53.42507154674088]:38.25619549,leaf\_1146[18.439048081586712]:3.270172027)node\_1152[15.168876054251774]:6.837512472)node\_1099[8.331363582073376]:2.633818457)node\_1039[5.6975451254480305]:1.549314134)node\_890[4.14823099105802]:0.7375491204)node\_511[3.4106818706794892]:0.8417615721)node\_97[2.568920298626249]:0.9004423404)node\_1[1.6684779582722558]:1.6684779582722558;</tree>  
 <epicontrib>0.5</epicontrib>  
 <changetimes>0</changetimes>  
</data>

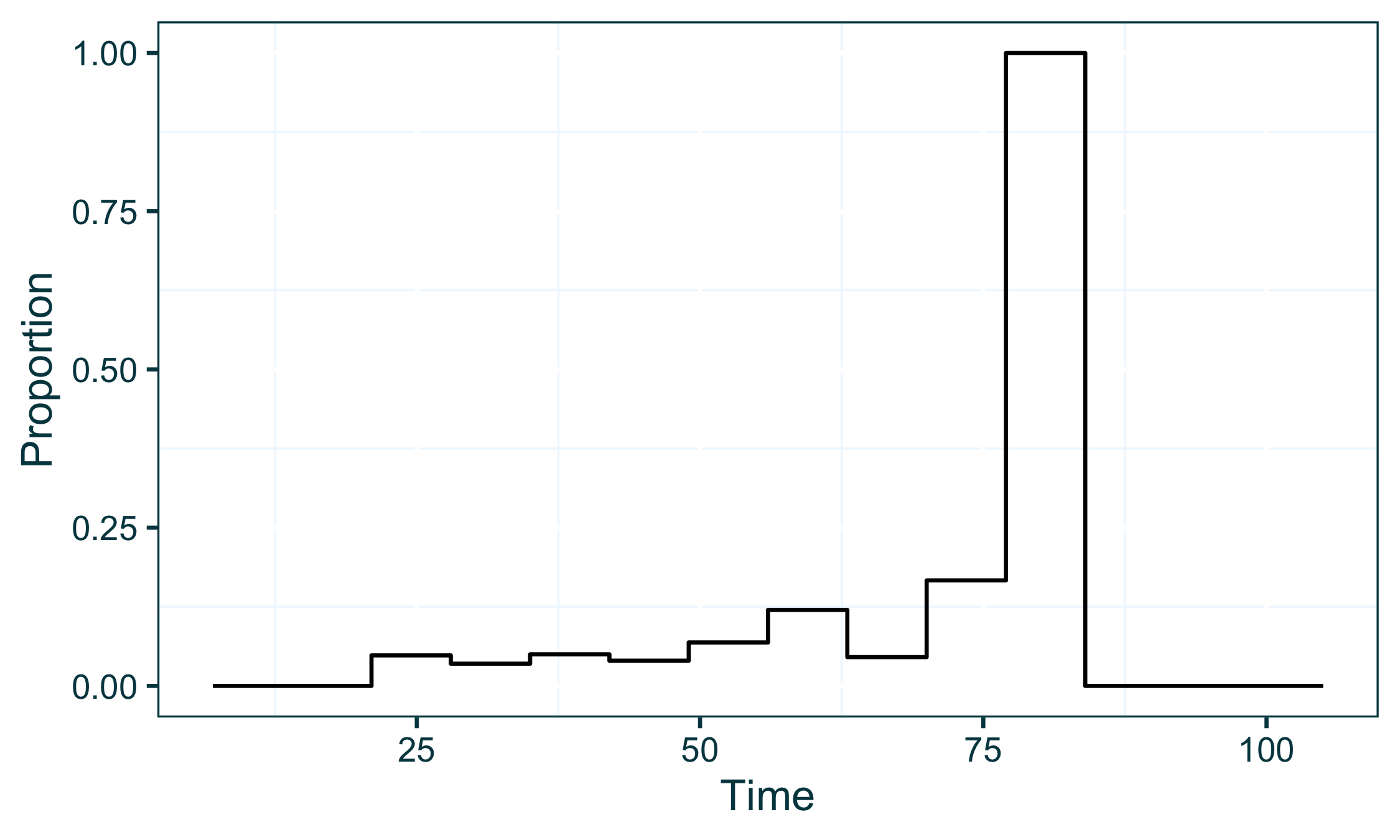
This data chunk inputs the below data into EpiFusion - weekly case incidence (totalling 1185 cases), and a phylogenetic tree of sequenced samples (59 sequences in total).

## Warning: The `size` argument of `element\_rect()` is deprecated as of ggplot2 3.4.0.  
## ℹ Please use the `linewidth` argument instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.



If we look at the number of sampled sequences (tree tips) over time as a proportion of the corresponding cases, we get something that looks like this:

## Joining with `by = join\_by(Time)`



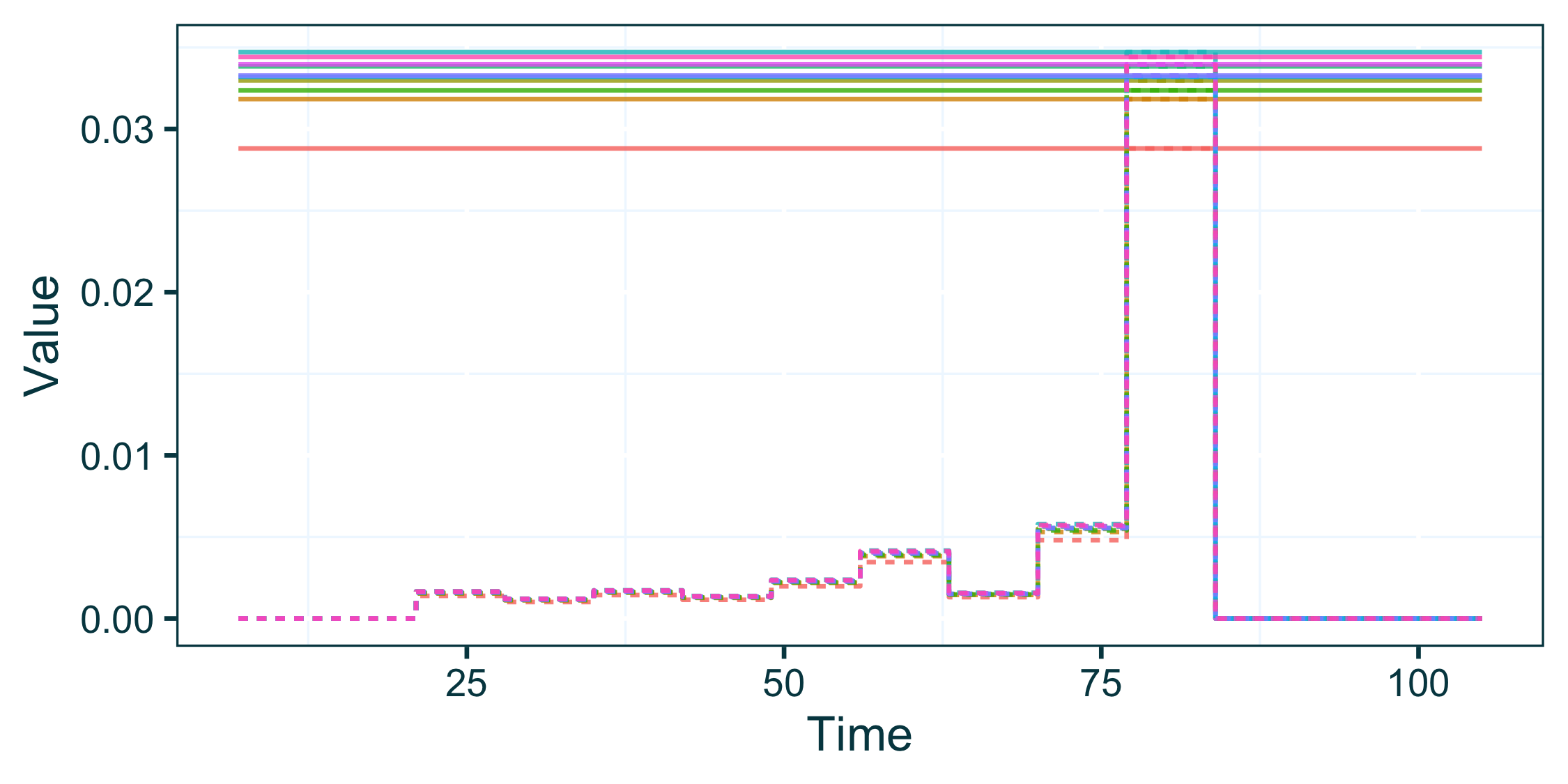
### Pairing the psi and phi parameters

In a situation like this where the user doesn’t want to provide priors for the psi parameter specifically, the XML file can be adjusted to contain the following in place of a <psi> block within the <priors> section. Below, the pairedPsi tag indicates that the user is not specifically parameterising psi, but instead it will be calculated using the parameterisation of the phi case sampling rate and the proportion of sequences to cases as which is calculable from the data.

<priors>  
 <gamma>  
 <!-- gamma parameterisation goes here -->  
 </gamma>  
 <pairedPsi></pairedPsi>  
 <phi>  
 <!-- phi parameterisation goes here -->  
 </phi>   
 <initialBeta>  
 <!-- initalBeta parameterisation goes here -->  
 </initialBeta>  
 <betaJitter>  
 <!-- betaJitter parameterisation goes here -->  
 </betaJitter>  
</priors>

### Effect in the MCMC sampling process

The resulting effect in the MCMC sampling process within EpiFusion is that the phi case sampling rate is fitted as a normal MCMC parameter, and the psi sequence sampling rate is fitted accordingly as a proportion of phi across time calculated from the data. This is demonstrated below, where 10 samples of the phi sampling rate and their corresponding psi values are plotted. Different colours indicate different MCMC samples, with the solid line showing phi and the dashed line showing the corresponding psi.

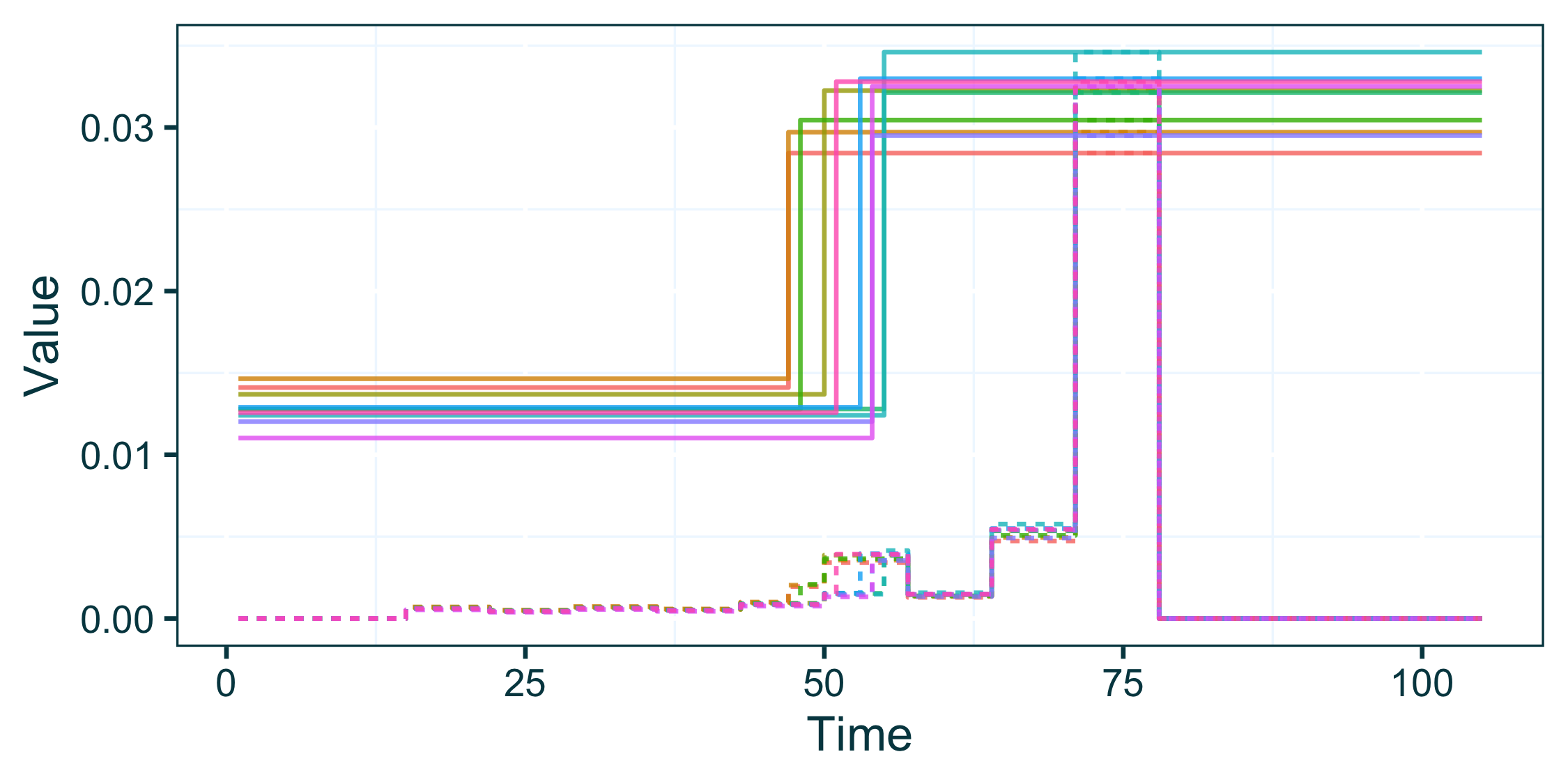


As is evident in the above graph, phi is parameterised as a constant rate over time (the solid horizontal lines). To demonstrate the paired psi feature wtih a more advanced parameterisation of phi, the following priors XML chunk can be used.

<priors>  
 <gamma>  
 <stepchange>false</stepchange>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.143</mean>  
 <standarddev>0.05</standarddev>  
 <lowerbound>0.0</lowerbound>  
 <upperbound>1.0</upperbound>  
 </gamma>  
 <pairedPsi></pairedPsi>  
 <phi>  
 <stepchange>true</stepchange>  
 <changetime>  
 <x0>  
 <disttype>UniformDiscrete</disttype>  
 <min>45</min>  
 <max>55</max>  
 </x0>  
 </changetime>  
 <distribs>  
 <x0>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.01</mean>  
 <standarddev>0.002</standarddev>  
 <lowerbound>0.0</lowerbound>  
 <upperbound>1.0</upperbound>  
 </x0>  
 <x1>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.03</mean>  
 <standarddev>0.01</standarddev>  
 <lowerbound>0.0</lowerbound>  
 <upperbound>1.0</upperbound>  
 </x1>  
 </distribs>  
 </phi>  
 <initialBeta>  
 <stepchange>false</stepchange>  
 <disttype>Uniform</disttype>  
 <min>0.2</min>  
 <max>0.9</max>  
 </initialBeta>  
 <betaJitter>  
 <stepchange>false</stepchange>  
 <disttype>Uniform</disttype>  
 <min>0.001</min>  
 <max>0.1</max>  
 </betaJitter>  
</priors>

The above amended XML chunk, specifically the phi block, parameterises a step-change in the phi case sampling rate at some point during the outbreak, with a uniform prior on the exact time of this change being between day 45 and 55. Some resulting phi samples, and paired psi, look like this.

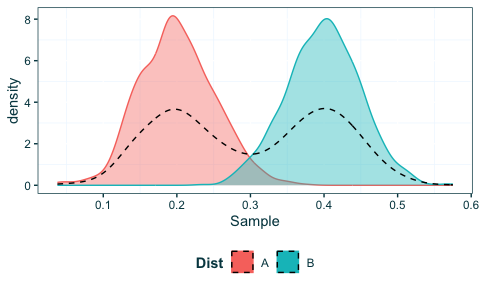
## Joining with `by = join\_by(Sample, Time)`



## Appendix 5 - Composite Prior Distributions

It is not uncommon to want to specify priors or distributions of variables that have a non-parametric form. To accomodate this we include the option to EpiFusion to specify priors with non-parametric distributions that are the composite of two or more parametric distributions

a <- data.frame(Dist = "A", Sample = rtruncnorm(1000, mean = 0.2, sd = 0.05, a = 0, b = 1))  
b <- data.frame(Dist = "B", Sample = rtruncnorm(1000, mean = 0.4, sd = 0.05, a = 0, b = 1))  
combined <- rbind(a, b)  
  
ggplot(combined, aes(x = Sample)) +  
 geom\_density(aes(col = Dist, fill = Dist), alpha = 0.4, bounds = c(0, 1)) +  
 geom\_density(linetype = 2, bounds = c(0, 1)) +  
 lshtm\_theme()



### Parameterisation in EpiFusion XML

In EpiFusion XML, prior distributions are specified inside the priors block, inside XML tags that describe the different variables (most essentially, gamma, psi, phi and some parameterisation of beta depending on the fitting method). For a given parameter, e.g. phi this can look a little like this:

<phi>  
 <stepchange>false</stepchange>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.2</mean>  
 <standarddev>0.05</standarddev>  
 <lowerbound>0.0</lowerbound>  
</phi>

In the above example, our parameterisation of phi is constant throughout the time series we are modelling (indicated by <stepchange>false</stepchange>). Our prior for phi is a simple truncated normal distribution.

To adjust this so that our prior is made of two composite distributions (like what is shown above) requires two small changes to the xml.

1. Add a new tag to the XML chunk numdists specifying the number of distributions involved
2. Enclose the details of the distributions into new ‘sublevel’ tags labelled with letters of the alphabet

In practice, here’s what the adjusted XML should look like:

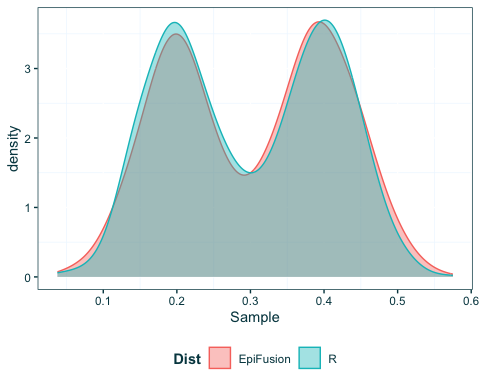
<phi>  
 <stepchange>false</stepchange>  
 <numdists>2</numdists>  
 <a>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.2</mean>  
 <standarddev>0.05</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </a>  
 <b>  
 <disttype>TruncatedNormal</disttype>  
 <mean>0.4</mean>  
 <standarddev>0.05</standarddev>  
 <lowerbound>0.0</lowerbound>  
 </b>  
</phi>

Above we specify that the phi prior is made up of two distributions, (<numdists>2</numdists>), which we have enclosed in chunks <a> and <b>.

### Validation by sampling from EpiFusion distribution

We generated 1000 samples from an EpiFusion prior distribution parameterised with the above XML code to verify that the expected distribution shape was obtained. Below we show these samples plotted with the samples generated for the above plot.

epifusion\_samples <- read.table("Extra\_Manuscript\_Assets/Composite\_Distribution\_Priors/composite\_distribution\_samples.txt", header = F) %>%  
 rename(Sample = V1) %>%  
 mutate(Dist = "EpiFusion")  
   
epifusion\_and\_r\_samples <- rbind(combined, epifusion\_samples) %>%  
 mutate(Dist = ifelse(Dist != "EpiFusion", "R", Dist))  
  
ggplot(epifusion\_and\_r\_samples, aes(x = Sample)) +  
 geom\_density(aes(col = Dist, fill = Dist), alpha = 0.4, bounds = c(0, 1)) +  
 lshtm\_theme()



### Calculating the Prior Probability from the Composite Distribution

To ensure that the prior probability of this composite distribution is accurately calculated within EpiFusion for the Metropolis Hastings Algorithm, we calculate below the ECDF of the distribution generated by R, and compare it to the values obtained from EpiFusion for the encoded composite distribution.

r\_cdf\_function <- ecdf(combined$Sample)  
r\_cdf <- data.frame(Dist = "R", Value = seq(0.01, 0.6, 0.01), Density = r\_cdf\_function(seq(0.01, 0.6, 0.01)))  
epifusion\_cdf <- data.frame(Dist = "EpiFusion", Value = seq(0.01, 0.59, 0.01), Density = read.table("Extra\_Manuscript\_Assets/Composite\_Distribution\_Priors/composite\_distribution\_density.txt", header = F)[,1]) %>%  
 mutate(Density = Density / sum(Density)) %>%  
 mutate(Density = cumsum(Density))  
  
epifusion\_and\_r\_cdf <- rbind(r\_cdf, epifusion\_cdf)  
  
ggplot(epifusion\_and\_r\_cdf, aes(x = Value, y = Density)) +  
 geom\_line(aes(col = Dist), alpha = 0.6) +  
 lshtm\_theme()

