Chapter 4
FoI and Age-Dependent Incidence

4.1 Burden of Disease

In everyday conversation about contagious maladies, "disease" and "infection" are sometimes used interchangeably. Often this imprecision does not matter. It is however useful to keep in mind that disease strictly speaking refers to symptomology and infection to pathogen/parasite colonization-status. The latent period —the time between a pathogen colonizes a host and the host can pass the infection on—is different from the incubation period—the time from colonization to onset of symptoms ("disease"). Such distinctions are obvious for certain infections; We all recognize the distinction between *AIDS* and *HIV positive*. The former refers to disease status, the latter to infection status. For "the flu," the virus is typically cleared in less than a week, but noncontagious cough and discomfort can last for another week or more. Thus clinical relevance is not always the same as dynamic relevance.

The severity of disease of many infections depends on age. The very young are often prone to more severe disease. Both measles and whooping cough, for example, cause highest morbidity and mortality in children under one (e.g., Miller and Fletcher 1976; Grais et al. 2007). Other diseases are more severe in the elderly. Mortality from influenza-like illness is a common example. "Teratogenic" diseases are those that cause complications during pregnancies. Rubella, chicken pox, and

This chapter uses the following R-packages: `splines` and `fields`.
A conceptual understanding of *Force of Infection* is useful prior to this discussion. A 5-min epidemics-MOOC intro can be watched from YouTube:
Force of Infection https://www.youtube.com/watch?v=dj1DiqA4Lvg.
Pathogens and Extinction https://www.youtube.com/watch?v=v67gtiACBTY.

O. N. Bjørnstad, *Epidemics*, Use R!, https://doi.org/10.1007/978-3-319-97487-3_4

Zika are important examples (Metcalf and Barrett 2016). For these, infections of reproductive-age women are the most pressing public health concern. It is important to understand determinants of age-prevalence curves for two reasons: First, because of such age-specificity in burden of disease, and second—as we shall see—because age-structure can mold infectious disease dynamics in important ways.

4.2 Force of Infection

The Force of Infection (FoI) is the *per capita* rate at which susceptibles are exposed to infection. The FoI in the S(E)IR compartmental model (Eqs. (2.1)–(2.3) and (3.1)–(3.3)) is $\phi = \beta I/N$ because each susceptible is assumed to contact other individuals in the population at some rate, the fraction of those contacts that are with infected individuals is I/N and β is by definition the contact rate times the probability of infection upon contact.[1]

An important basic and applied question is how the FoI scales with population density/size (de Jong et al. 1995). The literature suggests two extreme situations termed: "density-dependent" transmission for which the FoI scales linearly with density and "frequency-dependent" transmission for which the FoI is independent of density. Roberts and Heesterbeek (1993) points out that there is some significant confusion in the literature about the meaning of these terms, as the denominator N in the SEIR formulations is by some wrongly interpreted as Eqs. (3.3)–(3.5) being a "frequency-dependent" model. Roberts and Heesterbeek (1993) clarify that this is a mistaken interpretation; the I/N simply stems from the idea that only this fraction of random contacts are with infectious individuals (as opposed to the complimentary fraction which is with noninfectious individuals). The issue of density- *versus* frequency-dependence should be thought of in terms of how β (= contact rate $*$ transmission probability) scales with density (Roberts and Heesterbeek 1993; Ferrari et al. 2011). For the strictly density-dependent model, numbers of contacts are proportional to density, so $\beta(N) \propto N$ and thus transmission and R_0 scales linearly with density. In contrast the strictly frequency-dependent model assumes that contact rates are independent of N and, therefore, so is R_0. The frequency-dependent model is often used for sexually transmitted diseases (STDs) and vector-borne infections with the logic that the number of sexual partner does not scale with density and neither does the feeding requirements of mosquitos.

An interesting ecological implication is that in the absence of an alternative host, a deadly density-dependently transmitted pathogen is less likely to drive a host extinct because as the pathogen decimates the host, the reproductive ratio is expected

[1] The theoretical FoI is model specific, so more complicated models may have more complicated FoIs. The FoI for the SEIHFR model of Sect. 3.9.2, for example, is given by rate ① in Fig. 3.9.

to eventually decrease below one, at which time the chain-of-transmission will falter and break. Frequency-dependent pathogens, in contrast, may be able to sustain the chain-of-transmission to a bitter end as the reproductive ratio may remain supercritical (De Castro and Bolker 2005).

4.3 Probability of Infection at Age: The Catalytic Model

The FoI is a *rate*, thus if age-invariant, in a randomly mixing population the expected waiting-time to first infection is $1/\phi$. For endemic, fully immunizing infections in a constant-sized host population, R_0 determines the mean age of infection, $\bar{a}$, according to $R_0 \simeq 1 + L/\bar{a}$ where L is the life expectancy of the host. Thus the mean age of infection will be[2] $\bar{a} \simeq L/(R_0 - 1)$.

The general rate, $\phi(a,t)$, at which any susceptible will be infected may depend on age (a) and time (t). Ignoring time-dependence (but see Ferrari et al. 2010), the integrated rate of infection to age a is $\int_0^a \phi(a)da$, thus the probability of not being infected by age a is $1 - p(a) = exp(-\int_0^a \phi(a)da)$ and the probability of being infected on or before age a is (by the logic laid out in Sect. 3.2):

$$p(a) = 1 - e^{-\int_0^a \phi(a)da}. \tag{4.1}$$

This is called the catalytic model (Muench 1959; Hens et al. 2010).[3] Age-intensity curves and age-seroprevalence curves are important data-sources for estimating the FoI. For nonlethal, persistent infections and nonlethal, fully immunizing infections the former/latter provides excellent data for estimating ϕ. In the simplest case we assume that the FoI is independent of both age and time, in this case the probability of being infected by age a is $1 - \exp(-\phi a)$. If we have data on number of infected individuals by age, we can use the standard generalized linear model (`glm`) framework to estimate the FoI for this *simplest* model.

Generalized linear models have two components: an error distribution (such as binomial, Poisson, negative binomial, normal, etc.) and a "link" function which specifies how the expected (predicted) values $\hat{y}$ are linked to the "linear predictors" $x = a + b_1 x_1 + c_1 x_2 \cdots$. Common link functions are (depending on error distributions): "identity," "log," "logit" (= "log-odds" $= \log(\hat{y}/(1-\hat{y}))$), and "complimentary log-log" $(= \log(-\log(1-\hat{y})))$ (McCullagh and Nelder 1989). The link functions are associated with inverse link functions which for the aforementioned are: "identity," e^x, $\frac{e^x}{1+e^x}$, and $1 - e^{-x}$, respectively.

---

[2] In populations of changing size a more accurate calculation is $\bar{a} \simeq 1/(\mu(R_0 - 1))$, where μ is the host birth rate (Dietz and Schenzle 1985).

[3] If immunity wanes at a rate ω, the reversible catalytic model is $p(a) = \frac{\phi(a)}{\phi(a)+\omega}(1 - e^{-\int_0^a \phi(a)+\omega da})$ (see e.g., Pomeroy et al. 2015, for an example). Heisey et al. (2006) discuss corrections needed if infection causes significant disease-induced mortality.

Let us assume we test some n_a individuals of each age a and find from serology that i_a individuals have been previously infected. Inferring ϕ from this data is a standard(ish) binomial regression problem: $p(a) = 1 - exp(-\phi a)$ is the expected fraction infected (or seropositive) by age a. Thus $\log(-\log(1-p(a))) = \log(\phi) + \log(a)$, so we can estimate a constant log-FoI as the intercept from a `glm` with binomial error, a complimentary log-log link and log-age as a regression "offset."[4] The R call will be of the form[5]:

```
glm(cbind(inf, notinf) ~ offset(log(a)),
     family=binomial(link="cloglog"))
```

We can illustrate the approach using the pre-vaccination Measles antibody data of Black (1959). The data contain seroprevalence-by-age-bracket of some 300 people from around New Haven, Connecticut from blood drawn in the summer of 1957:

```
data(black)
black
```

```
##      age   mid  n pos neg         f
## 1     <1  0.75 10   8   2 0.8000000
## 2   1-4   2.50 21   4  17 0.1904762
## 3   5-9   7.00 41  31  10 0.7560976
## 4 10-14 12.00 52  50   2 0.9615385
## 5 15-19 17.00 30  28   2 0.9333333
## 6 20-29 25.00 38  37   1 0.9736842
## 7 30-39 35.00 51  49   2 0.9607843
## 8 40-49 45.00 35  31   4 0.8857143
## 9   >50 60.00 30  26   4 0.8666667
```

The age-profile of seroprevalence takes the characteristic shape of many pre-vaccination childhood diseases: High seroprevalence of the very young (<1 year) due to the presence of maternal antibodies that wanes with age, followed by rapid build-up of immunity to almost 100% seroprevalence by age 20 (Fig. 4.1). There is perhaps some evidence of loss of immunity in the elderly. We use the binomial regression scheme to estimate the log-FoI based on the data for people in the 1–40 year groups, and compare predicted and observed seroprevalence by age (Fig. 4.1):

[4] An offset is a covariate that has a fixed coefficient of unity in a regression.

[5] Binomial regression either takes a binary 0/1 variable as the response or a matrix with two columns representing number of successes and failures for each covariate level.

```
b2=black[-c(1,8,9),] #subsetting age brackets
#Estimate log-FoI
fit=glm(cbind(pos,neg) ~ offset(log(mid)),
    family=binomial(link="cloglog"), data=b2)
#Plot predicted and observed
phi=exp(coef(fit))
curve(1-exp(-phi*x), from=0, to=60,
     ylab='Seroprevalence', xlab='Age')
points(black$mid, black$f, pch='*', col='red')
points(b2$mid, b2$f, pch=8)
exp(fit$coef)
```

```
## (Intercept)
##   0.1653329
```

The estimated FoI is 0.16/year, giving a predicted mean age of infection of 6 years.

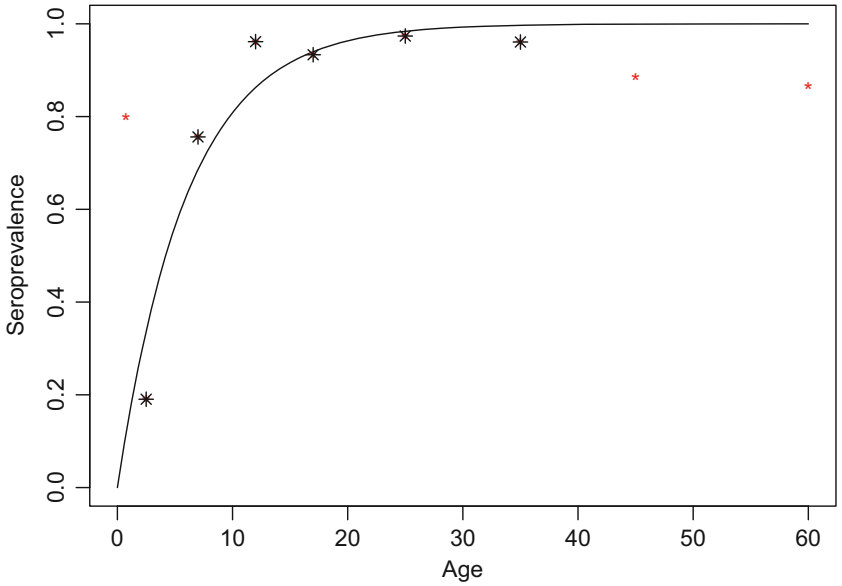

Fig. 4.1 Seroprevalence-by-age from the measles antibody study of Black (1959) from pre-vaccination Connecticut. The solid line is the predicted age-prevalence curve for the subset of the data used for estimation (black stars). The smaller red stars are data excluded from estimates due to maternal antibodies or possibly waning titers. Data are centered on the midpoints of each age-bracket

4.4 More Flexible ϕ-Functions

The assumption of a constant, age-invariant FoI is usually too simplistic because of age- or time-varying patterns of mixing. We can use Long et al.'s (2010) data on prevalence of the bacterium *Bordetella bronchiseptica* in a rabbit breeding facility to illustrate. *B. bronchiseptica* is a non-immunizing, largely avirulent (though it can cause snuffles), persistent infection of rabbits. Two-hundred-and-fourteen rabbits of known age were swabbed nasally and tested for the bacterium.

```
data(rabbit)
head(rabbit)
```

```
##     a  n inf
## 1 1.0 59   3
## 2 2.0  8   7
## 3 2.5  4   4
## 4 3.0  2   1
## 5 3.5  5   1
## 6 4.0  2   0
```

We first calculate the average FoI from the binomial regression scheme introduced above. In the breeding facility the older breeding animals are kept separate from the younger animals, so we restrict ourselves to rabbits <1 year old. We superimpose our fit on the plot of prevalence by age. In Fig. 4.2 the size of the circles is proportional to the sample size:

```
rabbit$notinf=rabbit$n-rabbit$inf
#Binomial regression
fit=glm(cbind(inf, notinf)~offset(log(a)),
    family=binomial(link="cloglog"),
    data=rabbit, subset=a<12)
#Plot data
symbols(rabbit$inf/rabbit$n~rabbit$a, circles=rabbit$n,
     inches=.5, xlab="Age", ylab="Prevalence")
#Predicted curves for <1 and all
phi=exp(coef(fit))
curve(1-exp(-phi*x), from=0, to=12, add=TRUE)
curve(1-exp(-phi*x), from=0, to=30, add=TRUE, lty=2)
1/phi
```

```
## (Intercept)
##    5.918273
```

The predicted median age of infection is just under 6 months. The constant-FoI model seems to do well for up to about 15 months of age, but the model overpredicts

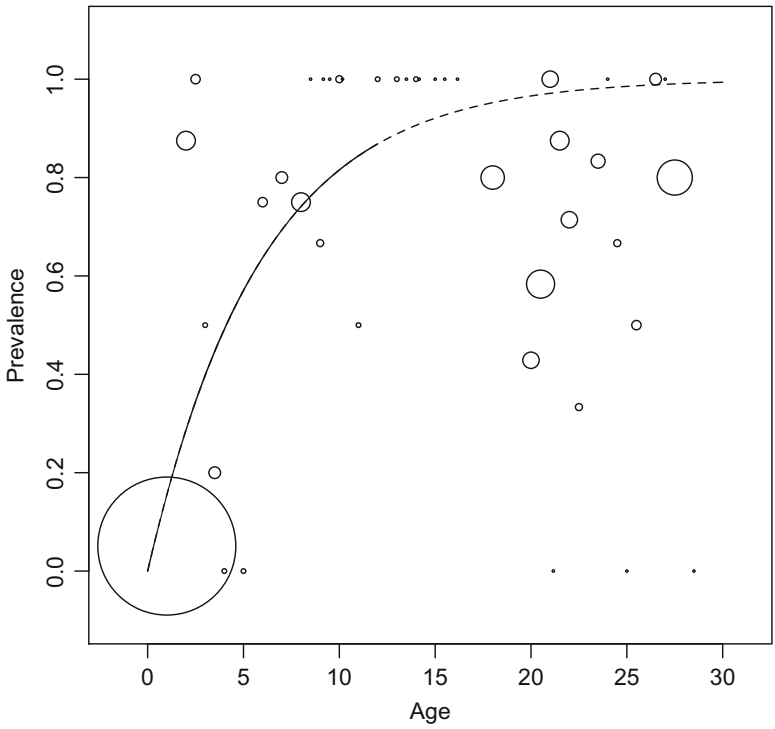

Fig. 4.2 Age-prevalence of *B. bronchiseptica* in a rabbit breeding facility. Circle size is proportionate to the number of animals tested in each age group. The solid line is the predicted age-prevalence curve for the subset of the data used for estimation (up to 1-year animals). The dotted line is the extrapolation to older individuals

the prevalence in older individuals. To allow for the scenario that the FoI varies with age, we need to implement our own framework (as opposed to using `glm`) using the maximum likelihood ideas introduced in Sect. 3.4. A simple model for age-specific FoI assumes a piecewise constant model (Grenfell and Anderson 1985), where individuals are classified into discrete age classes. For a piecewise constant model the integrand in Eq. (4.1) integrates to $\phi_a(a - c_a) + \sum_{k<a} \phi_k d_k$, where ϕ_a is the FoI of individuals in the a'th age bracket, and c_a and d_a are the lower cut-off age and duration of that bracket, respectively. We define a function for the integrand which takes the argument `a` for age, `up` is a vector of the upper cut-offs for each age bracket, and `foi` is the vector of age-specific FoIs:

```
integrandpc=function(a, up, foi){
  #Find which interval a belongs to
  wh=findInterval(a, sort(c(0,up)))
  #Calcultae duration of each interval
```

```
  dur=diff(sort(c(0,up)))
  #Evaluate integrand
  inte=ifelse(wh==1, foi[1]*a,
       sum(foi[1:(wh-1)]*dur[1:(wh-1)])+
          foi[wh]*(a-up[wh-1]))
  return(inte)
}
```

The negative log-likelihood function for the piecewise constant model takes arguments corresponding to log-FoI (`par`), age (`age`), number of positives (`num`), number tested in each age group (`denom`), and age-class cut-offs (`up`). Estimating the FoI on a log-scale (`foi=exp(par)`) ensures that all rates will be positive.

```
llik.pc = function(par, age, num, denom, up) {
    ll = 0
    for (i in 1:length(age)) {
       p = 1 - exp(-integrandpc(a=age[i], up = up,
          foi = exp(par)))
       ll = ll + dbinom(num[i], denom[i], p, log = T)
    }
return(-ll)
}
```

We use 1, 4, 8, 12, 18, 24, and 30 months as cut-off points for the age categories and assign arbitrary initial values of 0.1 for each piece of the FoI-function:

```
x = c(1, 4, 8, 12, 18, 24, 30)
para = rep(0.1, length(x))
```

For the analysis we use the `optim`-function to find maximum likelihood estimates:

```
est = optim(par=log(para),fn=llik.pc, age=rabbit$a,
     num=rabbit$inf,  denom=rabbit$n, up=x,
     method="Nelder-Mead", control=list(trace=2))
```

The maximum likelihood estimates for the log-FoI is given in `est$par`. The associated age-specific FoIs are:

```
round(exp(est$par), 6)
```

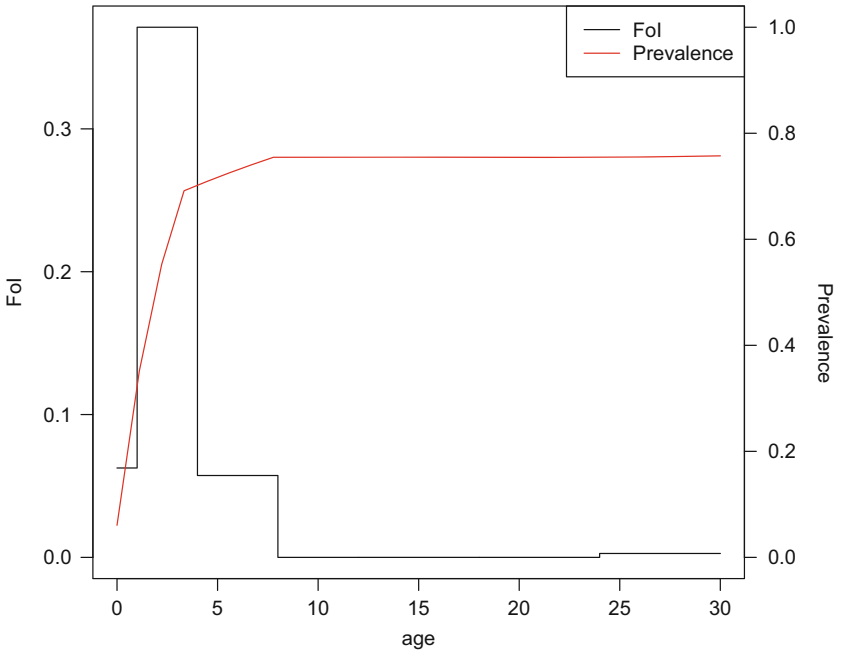

Fig. 4.3 The piecewise constant age-specific FoI of *B. bronchiseptica* in a rabbit breeding facility and the associated predicted age-prevalence curve

We can predict the age-prevalence curve and plot it as a step function (Fig. 4.3).

```
#Make space for left and right axes
par(mar = c(5,5,2,5))
#Add beginning and ends to x and y for step plot
xvals=c(0,x)
yvals=exp(c(est$par, est$par[7]))
plot(xvals, yvals, type="s", xlab="age", ylab="FoI")

#Superimpose predicted curve
par(new=T)
p = rep(0, 28)
for (i in 1:28) {
     p[i] = 1 - exp(-integrandpc(a=i, up = x,
        foi = exp(est$par)))
}
plot(p~c(1:28), ylim=c(0,1), type="l", col="red",
     axes=FALSE, xlab=NA, ylab=NA)

#Add right axis and legend
```

```
axis(side = 4)
mtext(side = 4, line = 4, "Prevalence")
legend("right", legend=c("FoI", "Prevalence"),
     lty=c(1,1), col=c("black", "red"))
```

The FoI peaks perinatally and then falls to zero after the 8-month age class. This is likely due to the older breeder females being housed separately and only having contact with their kittens. Long et al. (2010) used this (in combination with some other analyses; see Sect. 15.4) to conclude that most infections happen at a young age from infected mothers to their offspring and then among litter mates.

4.5 A Log-Spline Model

An alternative nonparametric approach to the piecewise constant model is to use smoothing splines. A spline is a smooth curve that can take an arbitrary shape except that it is constrained to be continuous and with continuous first and second derivatives (Härdle 1990; Hastie and Tibshirani 1990). The popularity of splines in nonparametric regression stems from its computational tractability; A spline can be fit by multiple regression on a set of "basis function"-decompositions of a covariate. The `gam` and `mgcv` packages offer automated ways to fit a variety of spline-variants to binomial data (and any other error distribution within the exponential family). Unfortunately, as with the case of the piecewise constant model, fitting the log-spline model is a bit more involved because of the integration step in Eq. (4.1). The `splines` package has functions to create various spline-bases that can be used with `lm`; `predict.lm` can predict values for the spline given regression coefficients.

The approach taken here is a bit cheeky in that it "hi-jacks" a spline-regression object created using the `bs`-spline basis functions in combination with `lm` and use `optim` to update/override the regression coefficients in the `lm`-object until a maximum likelihood solution is found. First we set the number of degrees-of-freedom for the spline. The `dl`-object will end up as the hi-jacked object for the age-specific FoI (Long et al. 2010).

```
require(splines)
# Degrees-of-freedom
df = 7
# Construct dummy lm-object
dl = lm(inf ~ bs(a, df), data = rabbit)
```

We write a `tmpfn`-function to predict the spline on a log-transformed scale to ensure that the force-of-infection (FoI) is strictly positive:

```
tmpfn = function(x, dl) {
    x = predict(dl, newdata = data.frame(a = x))
    exp(x)
}
```

The `tmpfn2`-function calculates the negative log-likelihood of the FoI as we did in the `foipc`-function above. In contrast to the piecewise constant model, the integrated splines do not have a closed form solution, so we use Rs inbuilt numerical integrator, `integrate`:

```
tmpfn2=function(par,data, df){
   #Dummy lm-object
   dl=lm(inf~bs(a,df), data=data)
   #Overwrite spline coefficients with new values
   dl$coefficients=par
   #Calculate log-likelihood
   ll=0
   for(i in 1:length(data$a)){
     p = 1 - exp(-integrate(tmpfn, 0, i, dl = dl)$value)
     ll=ll+dbinom(data$inf[i],data$n[i],p,log=T)
   }
 return(-ll)
 }
```

We use arbitrary initial values and minimize the negative log-likelihood using `optim`.

```
 para=rep(-1, df+1)
 dspline = optim(par=para, fn=tmpfn2, data=rabbit,
      df=df, method="Nelder-Mead", control=
      list(trace=2, maxit=2000))
```

We can plot the resultant maximum likelihood fits (Fig. 4.4).

```
par(mar = c(5,5,2,5)) #Room for two axes
#Overwrite dummy-objects coefficients with MLEs
dl$coefficients=dspline$par
#Age-prevalce plot
plot(tmpfn(rabbit$a,dl)~rabbit$a, type="l", ylab="FoI",
     xlab="Age (mos)", las=1)
#Overlay FoI
par(new=T)
p = rep(0, 28)
for (i in 1:28) {
     p[i] = 1 - exp(-integrate(tmpfn, 0, i,
          dl = dl)$value)
```

```
}
plot(p~c(1:28), ylim=c(0,1), type="l", col="red",
     axes=FALSE, xlab=NA, ylab=NA)
axis(side = 4, las=1)
mtext(side = 4, line = 4, "Prevalence")
legend("topright", legend=c("FoI", "Prevalence"),
     lty=c(1,1), col=c("black", "red"))
```

Both the piecewise and spline models show strong evidence of age-specificity in the FoI with a peak in transmission somewhere between 1 and 5 months of age, suggesting that circulation is mainly among the young and among littermates (Long et al. 2010). We revisit on this case study in Sect. 15.4.

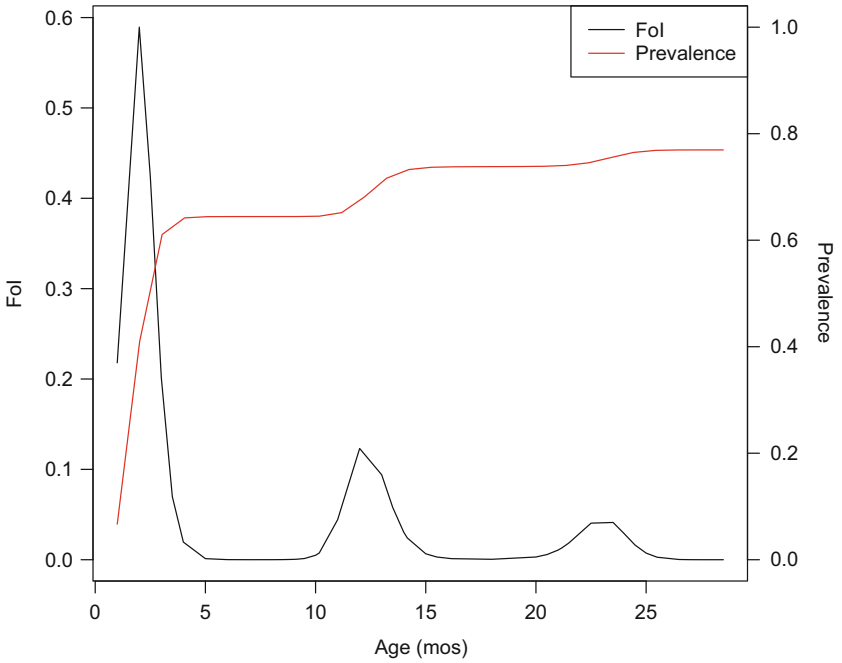

Fig. 4.4 The spline-estimate of the age-specific FoI of *B. bronchiseptica* in a rabbit breeding facility

4.6 Rubella

Rubella is a relatively mild, vaccine-preventable infection except that infection during pregnancy leads to stillbirths or congenital rubella syndrome. The main public health objective is therefore to minimize the FoI in women of childbearing age. The issue was made clear because of a surprising surge in CRS cases in Greece in the mid-90s following a low-intensity vaccination campaign (Panagiotopoulos et al. 1999).

Age-intensity data is less ideal than seroprevalence data for catalytic analysis; however, it is more common and therefore worth considering. Metcalf et al. (2011c) studied age-intensity curves for rubella across the provinces of Peru between 1997 and 2009. There were 24,116 reported cases during the period. The data are ½-monthly to age 1 and yearly thereafter (Fig. 4.5). With age-incidence data on immunizing infections, we can use the catalytic framework to estimate the relative age-specific FoI using the cumulative incidence by age (in place of age-seroprevalence or age-prevalence). For the analysis we use the total number of cases as our denominator because the actual number of susceptibles in each age group is not monitored. Hence, the estimate is a *relative* FoI because of the unknown baseline. Using the total cases as a denominator, further leads to sever biases of the FoI at old age classes (because exactly all of the assumed susceptibles in the final age class will be presumed to be infected at the time), so it should only be applied to the younger portion of the data. Its application also assumes a uniform age-distribution, so a correction for the age-pyramid may be necessary for a more refined analysis (Ferrari et al. 2010).

```
data(peru)
head(peru)
```

```
##           age incidence cumulative     n
## 2 0.01095890         1         56 24116
## 3 0.01369863         1         57 24116
## 4 0.01643836         1         58 24116
## 5 0.01917808         2         60 24116
## 6 0.03561644         1         61 24116
## 7 0.03835616         2         63 24116
```

```
#Calculate cumulative incidence
peru$cumulative=cumsum(peru$incidence)
#Define denominator
peru$n=sum(peru$incidence)
par(mar = c(5,5,2,5)) #Make room for two axes and plot
#Plot incidence with cumulative overlaid
plot(peru$incidence~peru$age, type="b", xlab="Age",
     ylab="Incidence")
par(new=T)
plot(peru$cumulative~peru$age, type="l", col="red",
```

```
     axes=FALSE, xlab=NA, ylab=NA)
axis(side = 4)
mtext(side = 4, line = 4, "Cumulative")
legend("right", legend=c("Incidence", "Cumulative"),
     lty=c(1,1), col=c("black", "red"))
```

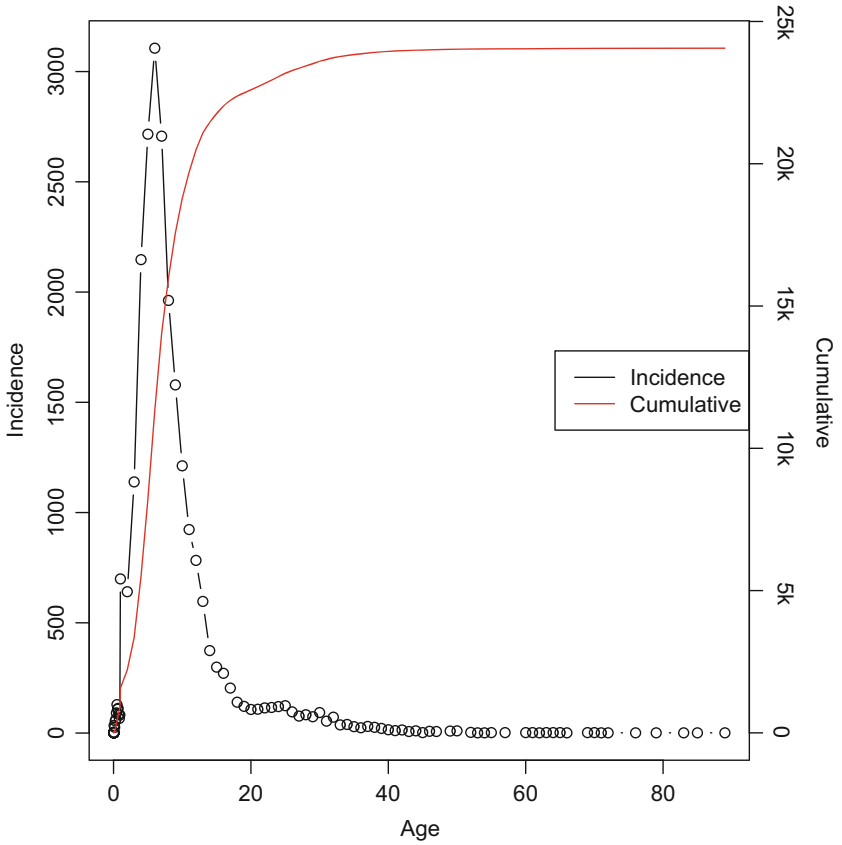

Fig. 4.5 Age-specific incidence and cumulative incidence of rubella in Peru 1997–2009

We first apply the piecewise model assuming a separate FoI for each year up to age 20 and 10 year classes thereafter. Convergence of the piecewise model with this many segments is very slow, so the actual figure (Fig. 4.6) was produced by doing repeat calls to `optim` using different optimization methods (`Nelder-Mead`, `BFGS`, and `SANN`), feeding the estimates from each call as starting values for the next. However, the basic analysis is:

```
#Upper age cut-offs
up=c(1:20,30, 40, 50, 60, 70,100)
para=rep(.1,length(up)) #Inital values
#Minimize log-likelihood
est2 = optim(par=log(para),fn=llik.pc, age=peru$age,
     num=peru$cumulative, denom=peru$n, up=up,
     method="Nelder-Mead", control=
     list(trace=2, maxit=2000))
#Step plot
x=c(0, up)
y=exp(c(est2$par, est2$par[26]))
plot(x, y, ylab="Relative FoI", xlab="Age", type="l",
     ylim=c(0,0.25), xlim=c(0,80))
```

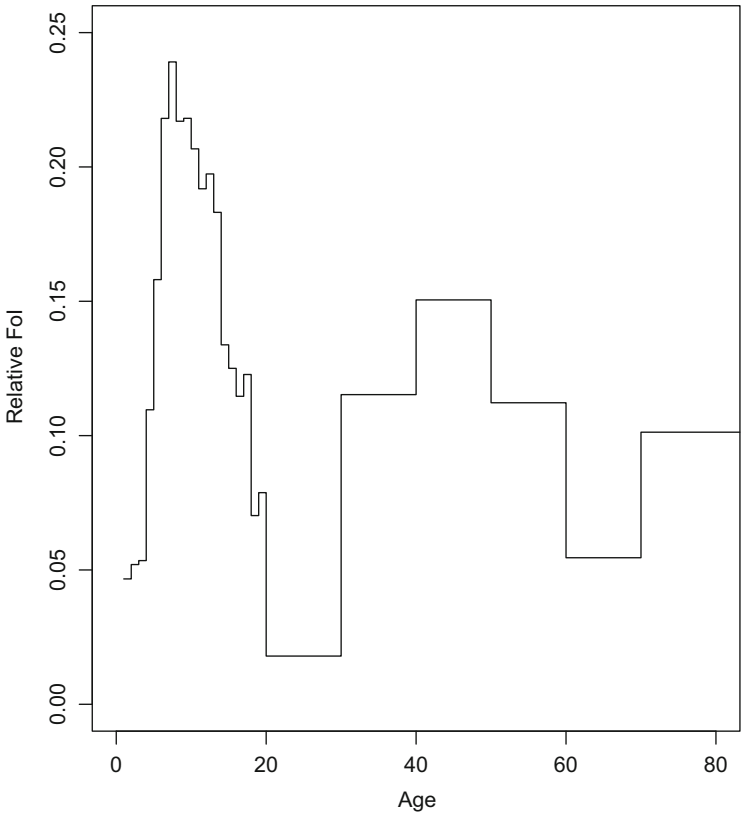

Fig. 4.6 The relative age-specific FoI of rubella in Peru as estimated using the piecewise-constant model

We see a clear peak in FoI in the 8–10 age group. The pattern makes sense given the biology of rubella and the assortative mixing commonly seen in the human host with most contacts being among same-aged individuals (see Sect. 4.7). Peru has a life-expectancy of around 75 years, and the R_0 of rubella is typically quoted in the 4–10 range, so according to $\bar{a} \simeq L/(R_0 - 1)$ the peak in circulation is predicted to be in an interval around 10 years of age.

We can do a more refined scenario-analyses regarding consequences of vaccination using the spline model. We focus on the 0–45-year age-range as this spans the pre to post child-bearing age:

```
data3 = peru[peru$age < 45, ]
df = 5
para = rep(0.1, df + 1)
```

We use a log-transformation to constrain the FoI to be positive, create the "dummy" `lm`-object, and define the function to evaluate the negative log-likelihood of the FoI curve given the data:

```
#Prediction function
tmpfn=function(x,dl){
    x=predict(dl, newdata=data.frame(age=x))
exp(x)}
#Dummy lm-object
dl=lm(cumulative~bs(age,df), data=data3)
#Log-likelihood function
tmpfn2=function(par,data, df){
    dl=lm(cumulative~bs(age,df), data=data)
    dl$coefficients=par
    ll=0
    for(a in 1:length(data$age)){
      p=((1-exp(-integrate(tmpfn,0,data$age[a],
         dl=dl)$value)))
      ll=ll+dbinom(data$cumulative[a],data$n[a],p,log=T)
    }
 return(-ll)
 }
```

Getting a good fit is, again, computationally expensive, but reveals an interesting two-peaked force-of-infection (Fig. 4.7): A dominant peak just under 10 years and a subdominant peak around 35. A plausible scenario is that most people get infected in school but the fraction that escapes this dominant mode of infection are most likely to contract the virus from their children when they reach school age.

```
#Fit model
dspline.a45.df5=optim(par=log(para),fn=tmpfn2,
     data=data3, df=df, method="Nelder-Mead",
```

```
     control=list(trace=4, maxit=5000))
#Overwrite dummy-objects coefficients with MLEs
dl$coefficients=dspline.a45.df5$par
plot(exp(predict(dl))~data3$age, xlab="Age",
     ylab="Relative FoI", type="l")
```

Fig. 4.7 The relative age-specific FoI of Rubella in Peru as estimated using the spline model

The fraction of cases that is predicted to occur in the child-bearing age-bracket (say, 15–40 years of age) is the joint probability of not being infected by age 15 and the probability of being infected in the 15–40 age range.

$$exp(-\int_0^{15} \phi(a)\,da)(1-exp(-\int_{15}^{40} \phi(a)\,da)) \tag{4.2}$$

We can predict this fraction from the spline model.

```
(exp(-integrate(tmpfn,0,15,dl=dl)$value))*(1-
     exp(-integrate(tmpfn,15,40,dl=dl)$value))

  ## [1] 0.08815273
```

Thus, with the current pattern of circulation just over 9% of the cases are predicted to occur in the at-risk age group. Let us ask how this fraction will change with a flat 50% reduction in FoI.

```
redn=0.5
(exp(-redn*integrate(tmpfn,0,15,dl=
     dl)$value))*(1-exp(-redn*integrate(tmpfn,
     15,40,dl=dl)$value))

  ## [1] 0.2376147
```

The reduction in FoI results, as predicted by theory, in an increase in the mean age of infection (in reality this will also likely lead to a change in the age-specific FoI curve), so that almost 24% of cases is predicted to fall in the at-risk group. Assuming an associated 50% reduction in cases, the total number in the age-bracket of concern would thus *increase* given this intervention—predicting an intervention-induced enhancement of the public health problem as was seen in Greece during the 1990s (Panagiotopoulos et al. 1999).

Metcalf and Barrett (2016) discuss public health issues related to the possible introduction of vaccines against Zika virus, which can cause microcephaly in children of mothers infected during pregnancy, in light of the lessons learnt from rubella. Whooping cough is another vaccine preventable disease that causes significant morbidity and mortality in perinatal children. Lavine et al. (2011) discuss how an imperfect (waning) vaccine could increase circulation among people of child-bearing age and thus increase the risk of parent-newborn transmission. They recommended that cocoon-vaccination of expecting parents should be considered if the current acellular vaccine is as leaky as is feared (Warfel et al. 2014). Althouse and Scarpino (2015) provide further discussion of the utility of cocoon-vaccination and other interventions.

4.7 WAIFW

Age-structured FoIs result from non-assortative mixing among different age groups. The Who-Acquires-Infection-From-Whom (WAIFW) matrix is used to describe the patterns of nonhomogenous mixing among different age groups (Grenfell and Anderson 1989). Mossong et al. (2008) conducted a diary-based social study to map age-stratified contact rates for various countries in Europe as part of the POLYMOD project. The contact rates by `contactor` and `contactee` are provided in the `mossong`-data set. We can visualize the diary data using an image plot with contours superimposed (Fig. 4.8)

```
data(mossong)
head(mossong)

  ##   contactor contactee contact.rate
  ## 1         1         1    120.37234
  ## 2         2         1     33.45833
  ## 3         3         1     23.13380
```

```
  ## 4          4          1      24.33333
  ## 5          5          1      29.00662
  ## 6          6          1      14.50331

x=y=mossong$contactor[1:30]
z=matrix(mossong$contact.rate, ncol=30, nrow=30)
image(x=x, y=y, z=z, xlab="Contactor",
     ylab="Contactee", col=gray((12:32)/32))
contour(x=x, y=y, z=z, add=TRUE)
```

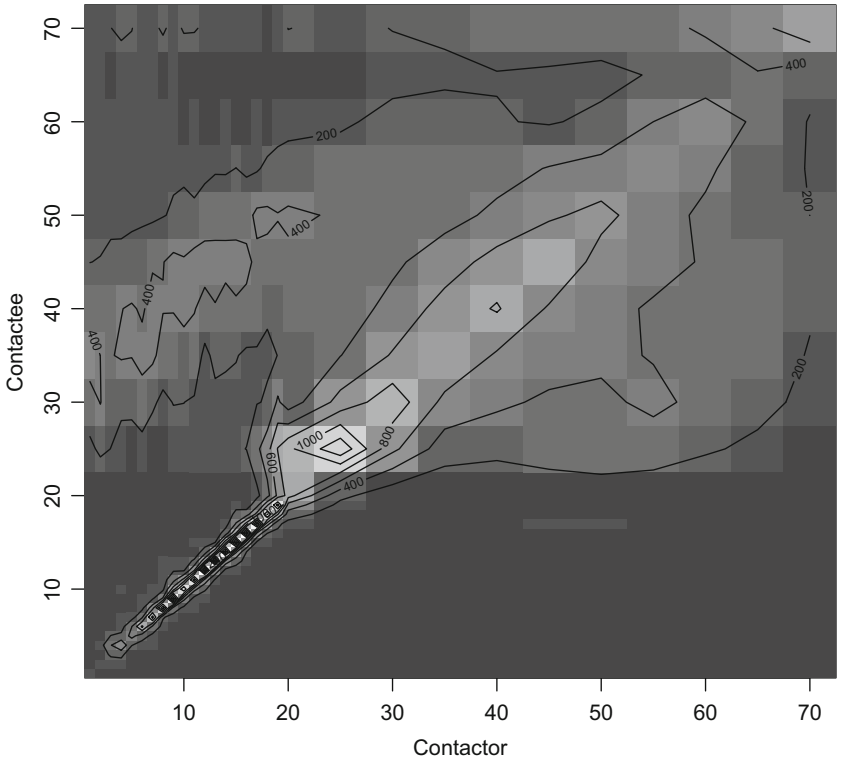

Fig. 4.8 The contact rates reported in the diary study of Mossong et al. (2008)

The reported contact rates are not symmetrical—which a WAIFW matrix will be—because of age-specific biases in diary entry rates as well as the age-profile of the contactors *versus* contactees. Before we "symmetrize" the matrix, we look at the reported marginal contact rate for each age group. Most contacts are among same-aged individuals and school-age children have the greatest number of contacts (Fig. 4.9). We do, however, also see off-diagonal ridges resulting from parent-offspring or children teacher interactions.

```
plot(apply(z,1,mean)~x, ylab="Total contact rate",
     xlab="Age")
```

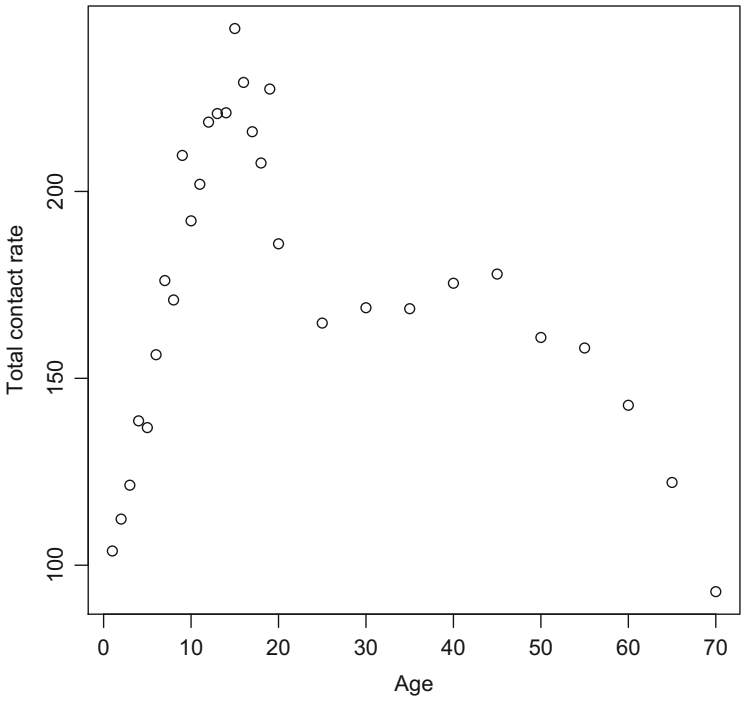

Fig. 4.9 The age-specific contact rates reported by the diary study of Mossong et al. (2008)

The symmetrized contact rate matrix (Fig. 4.10) is an estimate of the "WAIFW"-matrix.

4.8 Advanced: RAS Model

Schenzle (1984) discussed the importance of age-structured mixing when modeling infectious disease dynamics. Bolker and Grenfell (1993) extended this model to the "realistic age-structured (RAS) model" which in its full elaboration is an age-structured compartmental model with discrete aging of each birth cohort (at the beginning of each school year) and seasonality in transmission. Seasonality is the topic of Chap. 5. We can incorporate the POLYMOD contact matrix in a simpler age-structured model. We will make the simplifying assumptions that individuals age exponentially with rates set such that they will on average spend the right amount in each age-bracket. This allows us to formulate the model using chains of

ordinary differential equations. The upper-age cut-offs and age-progression rates for the $n = 30$ age categories are `x` and

```
a = c(1/diff(x), 0)
```

We can in principle use the raw symmetrized WAIFW matrix in our model, but we will use a thin-plate spline smoothed matrix using the `Tps`-function in the `fields`-package. The smoothing protocol also allows interpolation to use different age-brackets for the model than used in the contact survey whenever necessary (Fig. 4.10).

```
require("fields")
n=length(x)
z2=(z+t(z))/2
z3=as.vector(z2)
xy=data.frame(x=rep(x[1:n], n), y=rep(y[1:n], each=n))
polysmooth=Tps(xy, z3, df=100)
surface(polysmooth, xlab="", ylab="",
     col=gray((12:32)/32))

  ## [1] 6400    2
```

For our age-structured SIR model we first normalize the WAIFW matrix:

```
W=matrix(polysmooth$fitted.values[,
     1]/mean(polysmooth$fitted.values), nrow=n)
```

The age-specific force-of infection is $\phi = \beta \mathbf{W} \mathbf{I}/N$. The age-structured SIR model is thus (in log-coordinates)[6]:

```
siragemod = function(t, logx,  params){
     n=length(params$a)
     x = exp(logx)
     S = x[1:n]
     I = x[(n+1):(2*n)]
     R = x[(2*n+1):(3*n)]
     with(as.list(params), {
          phi = (beta*W%*%I)/N
          dS = c(mu,rep(0,n-1)) - (phi+a)*S +
              c(0,a[1:n-1]*S[1:n-1])*(1-p) - mu*S
          dI = phi*S + c(0,a[1:n-1]*I[1:n-1]) -
             (gamma+a)*I - mu*I
          dR =  c(0,a[1:n-1]*S[1:n-1])*p +
             c(0,a[1:n-1]*R[1:n-1]) + gamma*I -
```

[6] Recall that the `with(as.list(...))` allows us to evaluate the equations using the definitions in the `params`-vector.

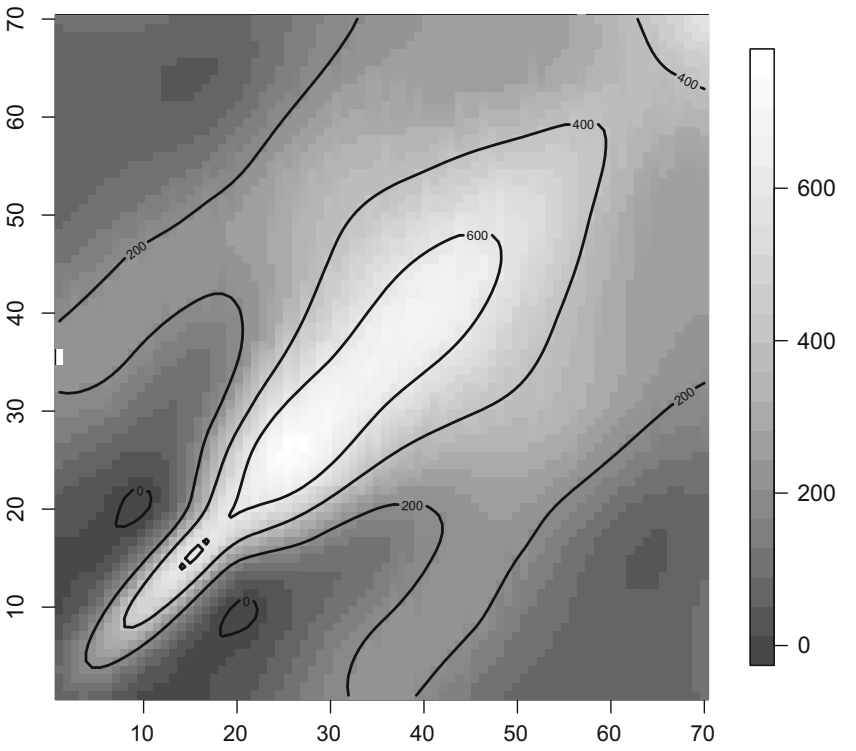

Fig. 4.10 The thin-plate spline smooth estimate of the WAIFW

```
            a*R - mu*R
        res = c(dS/S, dI/I, dR/R)
        list((res))
    })
}
```

where `S`, `I`, and `R` are vectors of length n, ϕ is the age-specific force of infection predicted by the WAIFW matrix, and p is a vector of length n that allows for age-specific vaccination rates (we will assume no vaccination). The a-vector sets appropriate aging rates when age groups vary in duration. We use the following parameters and initial conditions:

```
p.pre=rep(0,n)
pars.pre =list(N=1, gamma=365/14, mu=0.02, sigma=0.2,
     beta=100, W=W,p=p.pre, a=a)
ystart=log(c(S=rep(0.099/n,n), I=rep(0.001/n,n),
     R=rep(0.9/n,n)))
```

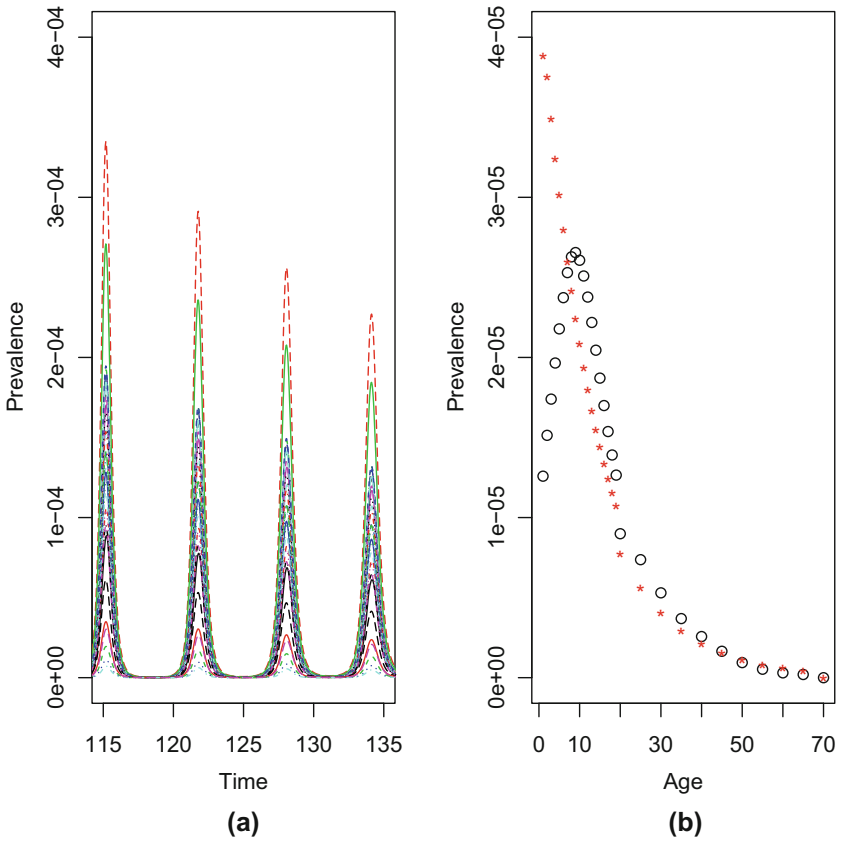

Fig. 4.11 The age-specific prevalences from the age-structured SIR model. (**a**) Trajectory through time. (**b**) Equilibrium age-incidence curves for the polymod matrix (o) vs homogenous mixing (∗)

and integrate to plot the age-specific I-dynamics (Fig. 4.11a) and equilibrium age-specific prevalence (Fig. 4.11b) for the polymod matrix. Figure 4.11b also shows the predicted age-prevalence curve for the age-structured model with homogenous mixing.

```
times=seq(0,500,by=14/365)
#Polymod mixing
out=as.data.frame(ode(ystart, times=times,
     func=siragemod, parms=pars.pre))
par(mfrow=c(1,2)) #Room for side-by-side plots
#Time series
matplot(times, exp(out[,32:61]), type="l", xlab="Time",
    ylab="Prevalence", xlim=c(50,90), ylim=c(0, 0.0005))
#Final age-prevalence curve
plot(x, t(exp(out[13036,32:61])*a), ylab="Prevalence",
```

```
    xlab="Age", ylim=c(0, 4E-5))
#Homogenous mixing:
pars.pre$W=matrix(1, ncol=30, nrow=30)
out2=as.data.frame(ode(ystart, times=times,
    func=siragemod, parms=pars.pre))
points(x, t(exp(out2[13036,32:61])*a), col=2, pch="*")
```

In contrast to the model with homogenous mixing which predicts that age-intensity curves decay exponentially with age, the RAS model can lead to a variety of age-incidence curves including the hump-shaped curve with a mode at around 10 years seen in Fig. 4.11b.