Harbin plague epidemic

February 15, 2017@ 8:19

It occurred to me that it would be interesting to contrast information that we have on 20th-century plagues (Bombay, 1906, bubonic; Harbin, 1911, pneumonic; others??) with the 14th- and 17th-century London data that David Earn has been collecting.

```
## Error in opts_chunk$set(fig.width = s, fig.height = 5, tidy = FALSE,
echo = FALSE, :  object 's' not found
```

Load packages:

```
library(deSolve)

## Warning:  package 'deSolve' was built under R version 3.3.2

library(ggplot2); theme_set(theme_bw())

## Warning:  package 'ggplot2' was built under R version 3.3.2

library(bbmle)
library(fitsir)
library(dplyr)
library(tidyr)
## if necessary:
## devtools::install_github("bbolker/fitsir")
```

From Dietz (2009) . . .

Figure 1 shows Dietz's plot – the only reference he gives to the data is "(International Plague Conference, 1912)" [not otherwise referenced in the paper!] Googling `'"international plague conference" harbin 1912'` does bring up some promising hits, especially this page, and particularly this PDF file, and particularly p. 529 of that page (Figure 2)

I used `g3data` to extract data points from Dietz's figure (before I found the 1912 report).

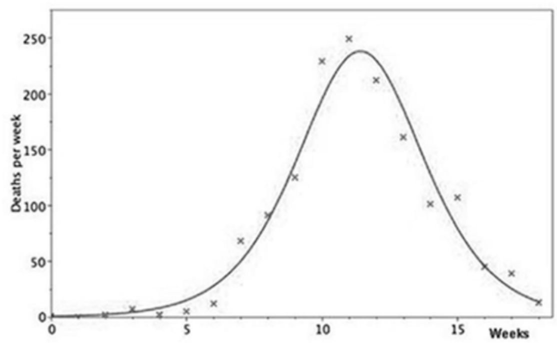


Figure 1: Unnumbered figure (p. 102) from Dietz (2009) showing the Harbin epidemic.

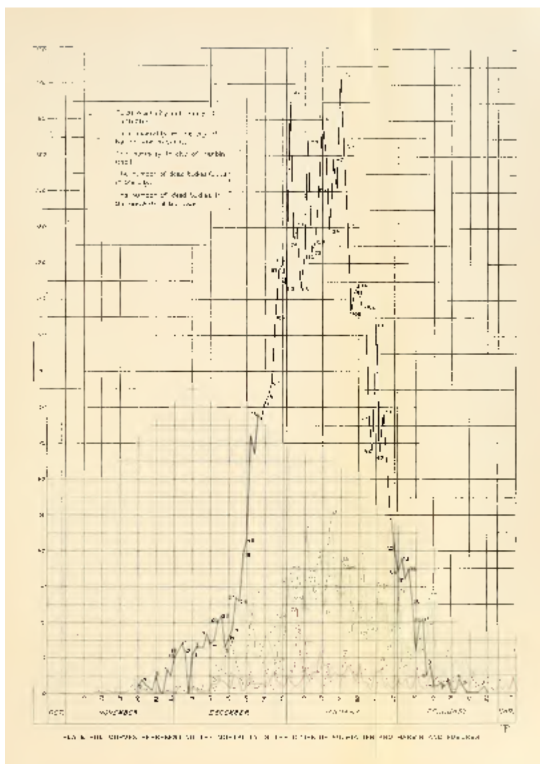


Figure 2: A thumbnail of the relevant page from International Plague Conference (1911 : Mukden) et al. (1912), extracted from the PDF via `pdftk A=reportofinternatinte.pdf cat A529-529 harbin_plague.pdf` ...

```r
dat <- read.csv("Dietz_harbin_sm.csv",header=FALSE)
names(dat) <- c("week","Deaths")
g0 <- ggplot(dat,aes(week,Deaths))+geom_point()+
    ## geom_smooth(method="loess")+
    ## geom_smooth(method="gam",method.args=list(family=quasipoisson),
    ##             formula=y~s(x,k=10),colour="purple")+
  coord_cartesian(ylim=c(0,300))
```

Dietz gives the (Kermack-McKendrick) equations for the incidence, dz/dt (based on a second-order Taylor expansion):

$$
\begin{aligned}
\frac{dz}{dt} &= \frac{\gamma x_0}{2\mathcal{R}_0^2} c_1 \mathrm{sech}^2(c_1 \gamma t - c_2), \\
c_1 &= \sqrt{(\mathcal{R}_0 - 1)^2 + \frac{2\mathcal{R}_0^2}{x_0}} \\
c_2 &= \tanh^{-1}\left(\frac{\mathcal{R}_0 - 1}{c_1}\right).
\end{aligned}
\tag{1}
$$

and estimates "$x_0 = 2985$, $\mathcal{R}_0 = 2.00$ and a mean infectious period of 11 days".

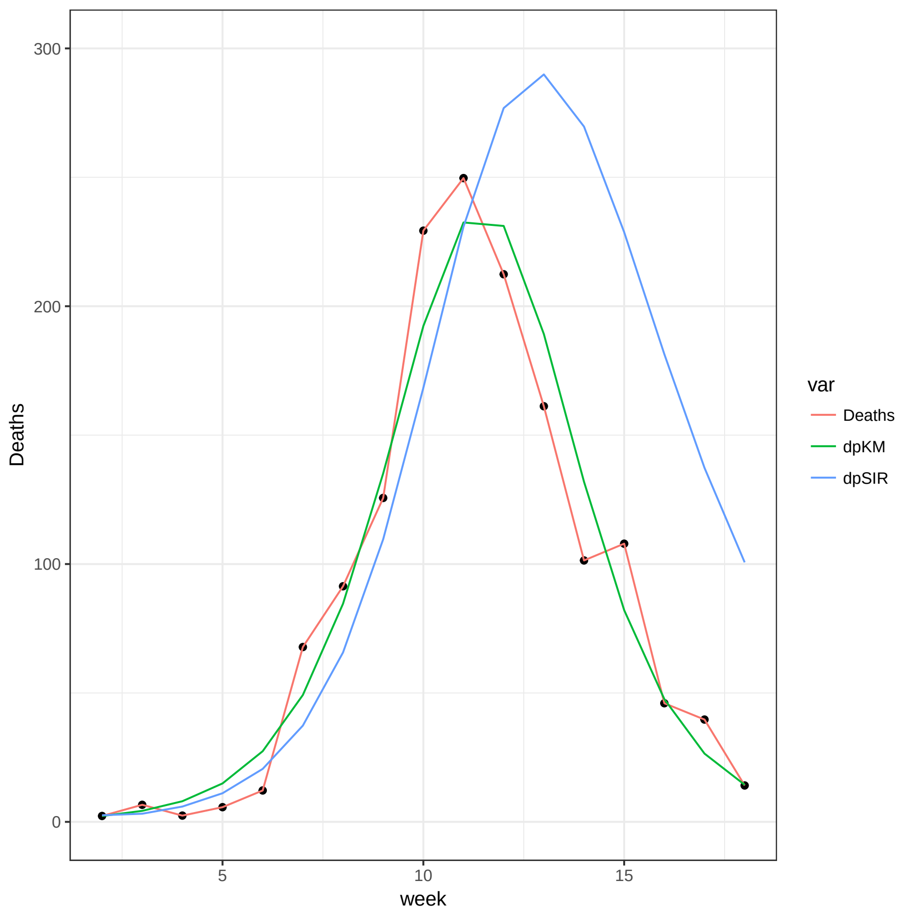
The weekly deaths should be approximately proportional to the incidence (this ignores the probability of survival, the integration over weeks, the second-order expansion, and all the other unrealities of the model . . .)

```r
dietz_harbin <- c(x0=2985,rzero=2,gamma=7/11)
gSIR <- function(t,y,params) {
  g <- with(as.list(c(y,params)),
        {
            ## R0 = beta*N/gamma
            beta <- rzero*gamma/x0
            c(S=-beta*S*I,
              I=beta*S*I-gamma*I,
              R=gamma*I)
        })
  list(g,NULL)
}
S0 <- c(S=unname(dietz_harbin["x0"])-1,I=1,R=0)
hfit1 <- ode(y=S0,
             times=c(0,dat$week),
             func=gSIR,
             parms=dietz_harbin)
dat$dpSIR <- diff(hfit1[,"R"])  ## pretend that incidence is per week
dat$dpKM <- with(as.list(dietz_harbin),
        {
           c1 <- sqrt((rzero-1)^2+2*rzero^2/x0) ## I think this is missing a term...
           c2 <- atanh((rzero-1)/c1)
```

```
            gamma*x0/(2*rzero^2)*c1*
                (1/cosh(c1*gamma*dat$week/2-c2))^2
        })
mdat <- dat %>%
    gather(var,val,-week)
```

Dietz gives a wrong equation!

```
g0 + geom_line(data=mdat,aes(x=week,y=val,colour=var))
```

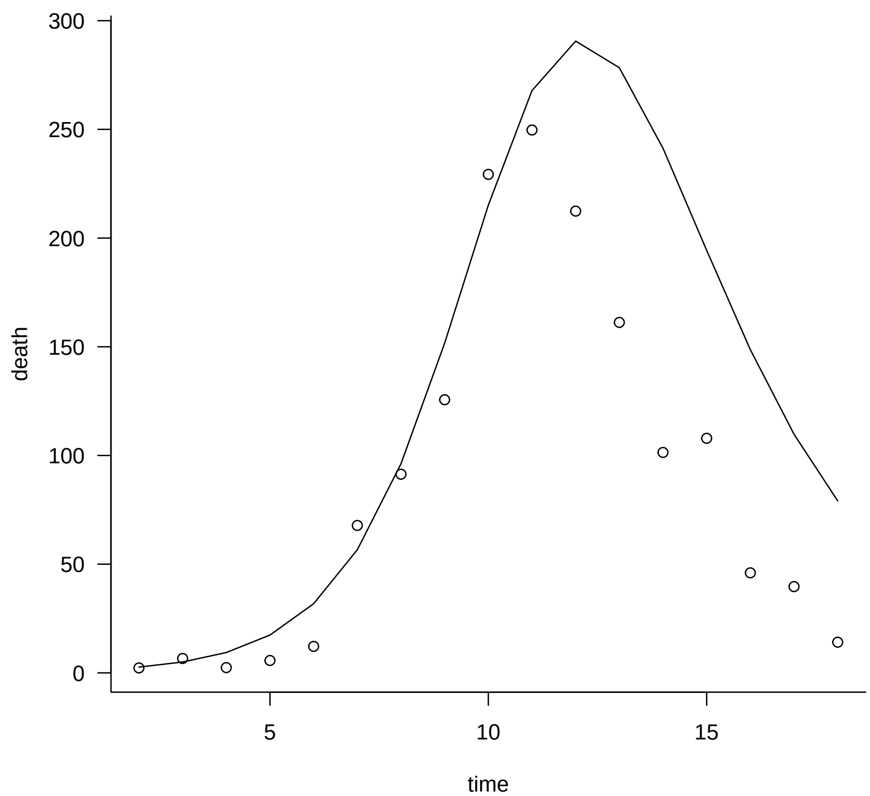


1 fitsir

Clearly, SIR model doesn't work very well!

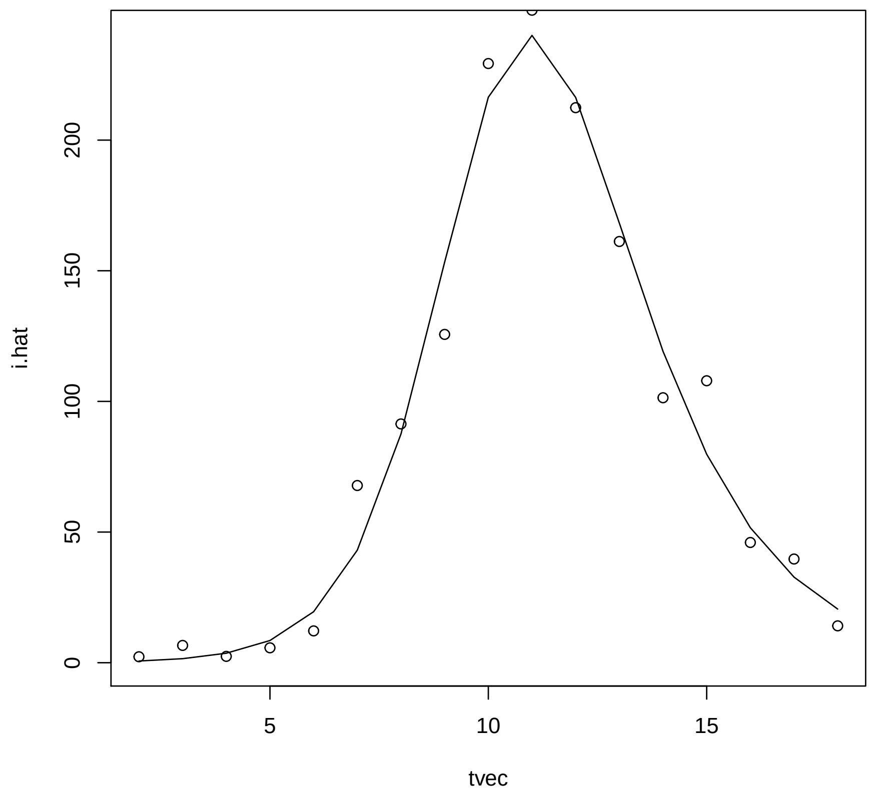
```r
dat1 <- setNames(dat,c("tvec","count"))
dietz_pars <- c(R0=2,gamma=7/11,N=2985)
dietz_lpars <- with(as.list(dietz_pars),
      c(log.beta=log(R0*gamma),
        log.gamma=log(gamma),
        log.N=log(N),
        logit.i=qlogis(1e-3)))
tvec <- dat1$tvec

par(las=1,bty="l")
plot(tvec, SIR.detsim(tvec, trans.pars(dietz_lpars), type = "death"), type = "l",
     xlab="time",ylab="death")
points(dat1)
```



Let's try fitting

```r
ff2 <- fitsir(dat1,start=dietz_lpars, type = "death")
plot(ff2)
```



It's similar to the parameters that Dietz give but it's not close enough. Why are they different?

KM assume that $\mathcal{R}_0 R/N$ is small.

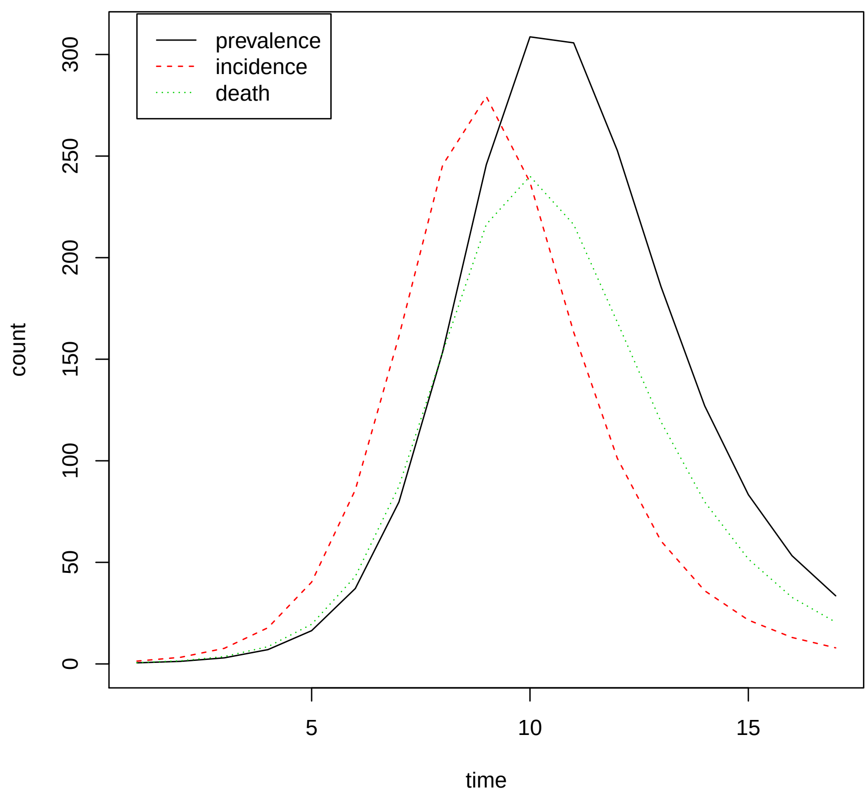
```r
summarize.pars(coef(ff2))

##           R0            r       infper           i0           I0
## 2.123742e+00 8.625436e-01 1.302823e+00 2.962312e-04 5.354921e-01
##           S0            N
## 1.807147e+03 1.807683e+03
```

We can also compare prevalence vs incidence vs deaths using this parameter:

```
fpars <- coef(ff2)
tpars <- trans.pars(fpars)
ss.p <- SIR.detsim(tvec, tpars)
ss.i <- SIR.detsim(tvec, tpars, type = "incidence")
ss.d <- SIR.detsim(tvec, tpars, type = "death")

matplot(data.frame(ss.p,ss.i,ss.d),type = "l",xlab="time",ylab="count")
legend(x=1,y=320,col=1:3,lty=1:3,legend=c("prevalence","incidence","death"))
```

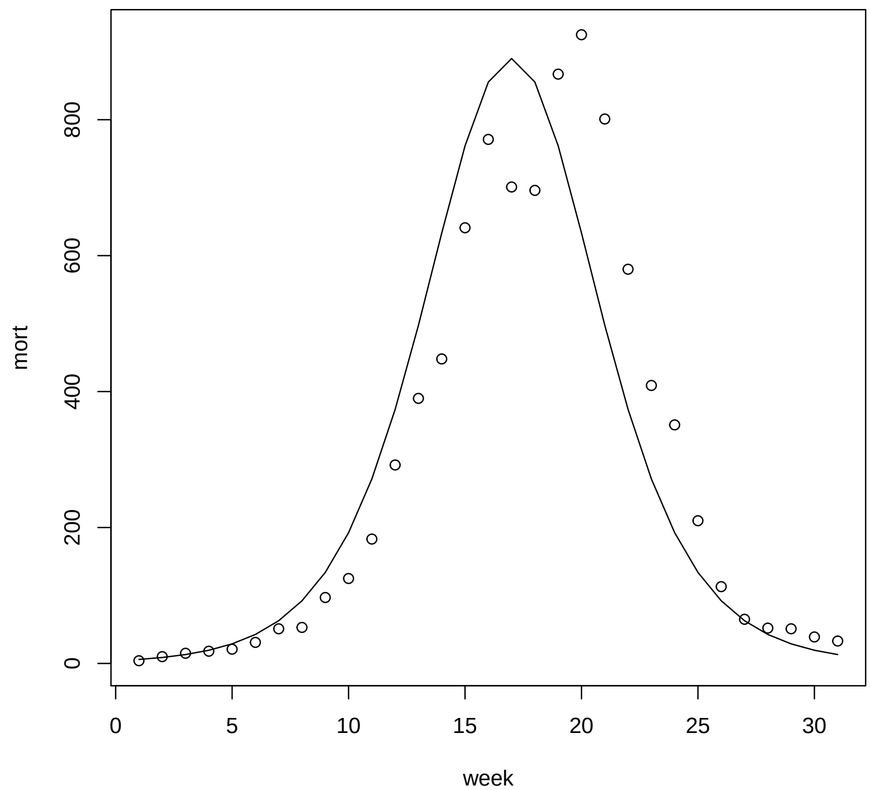


2 Bombay

We can use the Kermack and McKendrick equation to fit a curve to the bombay data. Equation that Kermack and McKendrick gives is

$$\frac{dz}{dt} = 890^2(0.2t - 3.4)$$

```r
plot(bombay)
lines(890*1/(cosh(0.2*1:31-3.4))^2)
```



We're able to reproduce their result. I think these parameters are close to their parameters:

```r
bpars <- c(log.beta = log(0.82),
           log.gamma = log(0.33),
           log.N = log(11300),
           logit.i = -8)

convertKM <- function(param){
    with(as.list(summarize.pars(param)),{
        c1 <- sqrt((R0-1)^2+2*R0^2/N)
        c2 <- atanh((R0-1)/c1)
        c(a=1/infper*N/(2*R0)*c1,
          b=1/infper*c1/2,
```

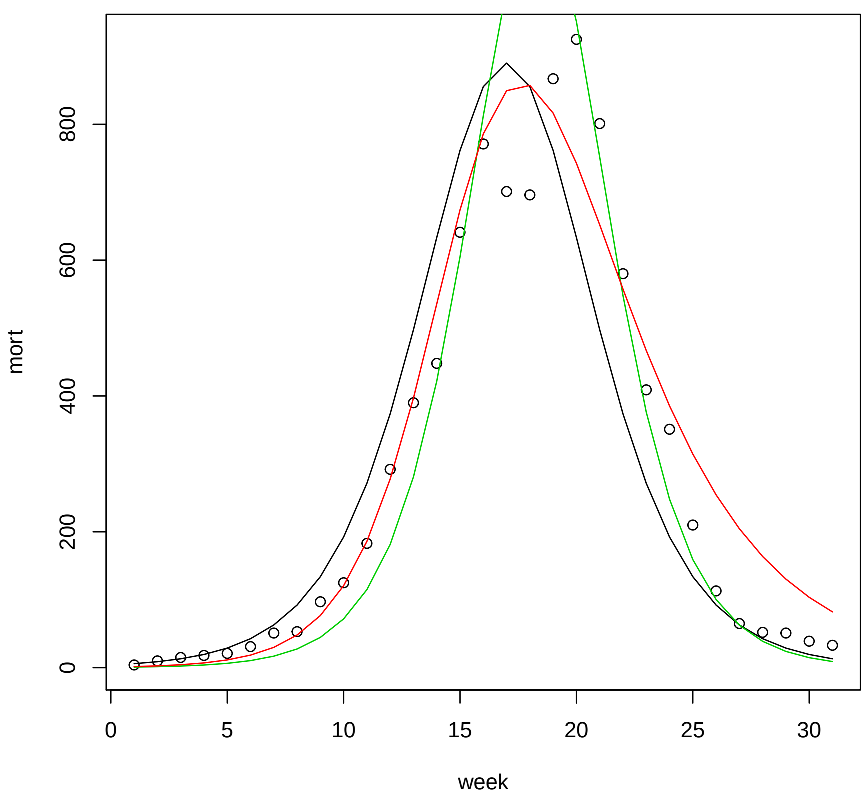
```r
           c=c2)
     })
}

convertKM(bpars)

##              a              b              c
## 1114.4285273    0.2450607    4.4980775

plot(bombay)
lines(890*1/(cosh(0.2*1:31-3.4))^2)
with(as.list(convertKM(bpars)),{
    lines(a*1/(cosh(b*1:31-c)^2), col =3)
})

lines(SIR.detsim(1:31, trans.pars(bpars), type = "death"), col = 2)
```



Now, let's try fitting an SIR model to the data:

```
bombay2 <- setNames(bombay, c("tvec", "count"))
## grad = TRUE because it's faster...
## we get stuck at a different local minima if we use NM
bfit <- fitsir(bombay2, start = bpars, type = "death")
```

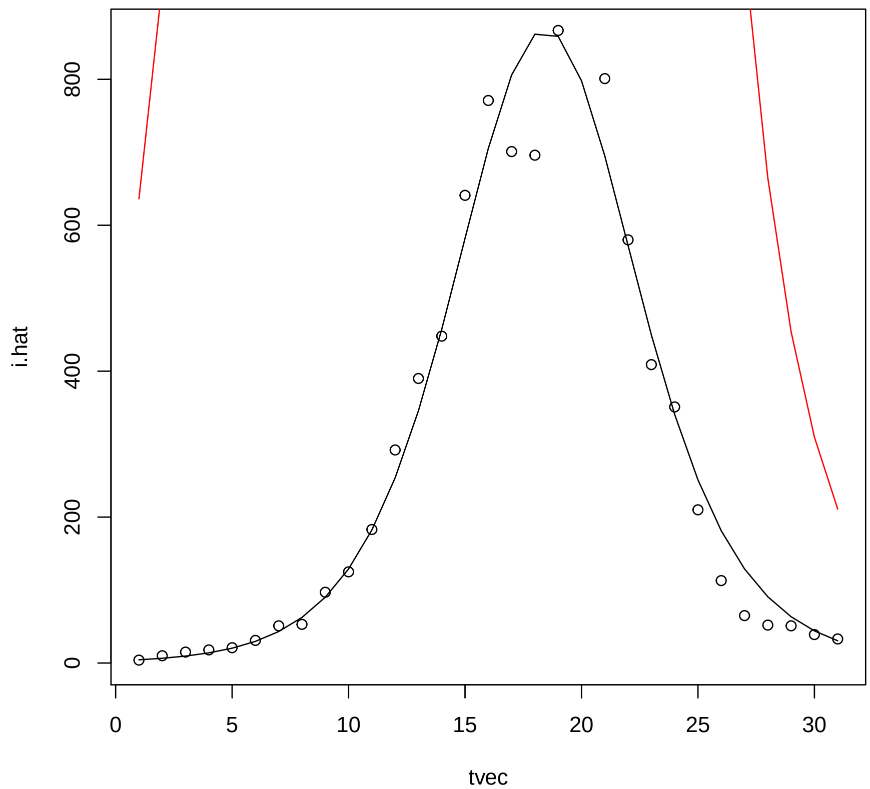
Let's look at the parameters

```
KMpars <- convertKM(coef(bfit))

print(KMpars)

##            a            b            c
## 29712.520752     0.192482     2.802057

plot(bfit)
with(as.list(KMpars),{
    lines(a*1/(cosh(b*1:31-c)^2), col =2)
})
```



References

Dietz, K. (2009, April). Epidemics: the fitting of the first dynamic models to data. *Journal of Contemporary Mathematical Analysis 44*(2), 97–104.

International Plague Conference (1911 : Mukden), R. P. R. P. Strong, G. F. Petrie, A. S. Megaw, and Boston College Libraries (1912). *Report of the International plague conference held at Mukden, April, 1911*. Manila, Bureau of Printing.