eAppendix

1. Stochastic SEIR Transmission Model Implementation

A sample outbreak is initialized by creating 153 households, with sizes h_i, drawn from the census distribution of household sizes. The initial household state is set to $q_{i,0} = \{(h_i - 1),0,1,0\}$, indicating that only the index case is symptomatic, all other household members being susceptible. The transmission model is summarized in the algorithm below, where S, E, I and R are the number of individuals in each state and the model is initialized at t=0:

```
If E + I > 0:
    For s in S:
        Draw x from Uniform(0,1]
        If x <= 1 − exp(−(βI + α)dt):
            S = S − 1
            E = E + 1
            Draw symptom onset time from Gamma(1/ε, ε_S)
            Draw recovery time from Gamma(1/γ, γ_S)
    t = t + dt

At end of step, transition from E → I and I → R those who have
symptom onset or recovery time <= t
```

eAlgorithm 1

The model is stepped forward in hourly increments ($dt = 1/24$), which gives a reasonable approximation of a continuous time infection process. Rates are expressed in terms of days but scaled to the appropriate time step.

The incubation and infectious periods are conceptualized as a sequence of e_s and i_s second-order compartments, with the probability of transition between these compartments for each individual equal to $(\varepsilon \cdot \varepsilon_s)dt$ and $(\gamma \cdot \gamma_s)dt$. This process yields $E \rightarrow I$ and $I \rightarrow R$ transition rates that are gamma distributed with means e, g and shape

parameters e_s, g_s, respectively. Transmission rates are also scaled in terms of dt (see Equation 1).

2. Asymptomatic Infections

To assess the effect of unobserved asymptomatic infections, we implemented the stochastic SEIR model outlined above, with an additional parameter, τ, that controls the proportion of new infections that are asymptomatic:

```
If E + I > 0:
    For s in S:
        Draw x from Uniform(0,1]
        If x <= 1 − exp(−(βI + α)dt):
            Draw y from Uniform(0,1]
            If y <= τ:
                S = S − 1
                R = R + 1
            Else:
                S = S − 1
                E = E + 1
                Draw symptom onset time from Gamma(1/ε, ε_S)
                Draw recovery time from Gamma(1/γ, γ_S)

    t = t + dt

At end of step, transition from E → I and I → R those who have symptom
onset or recovery time <= t
```

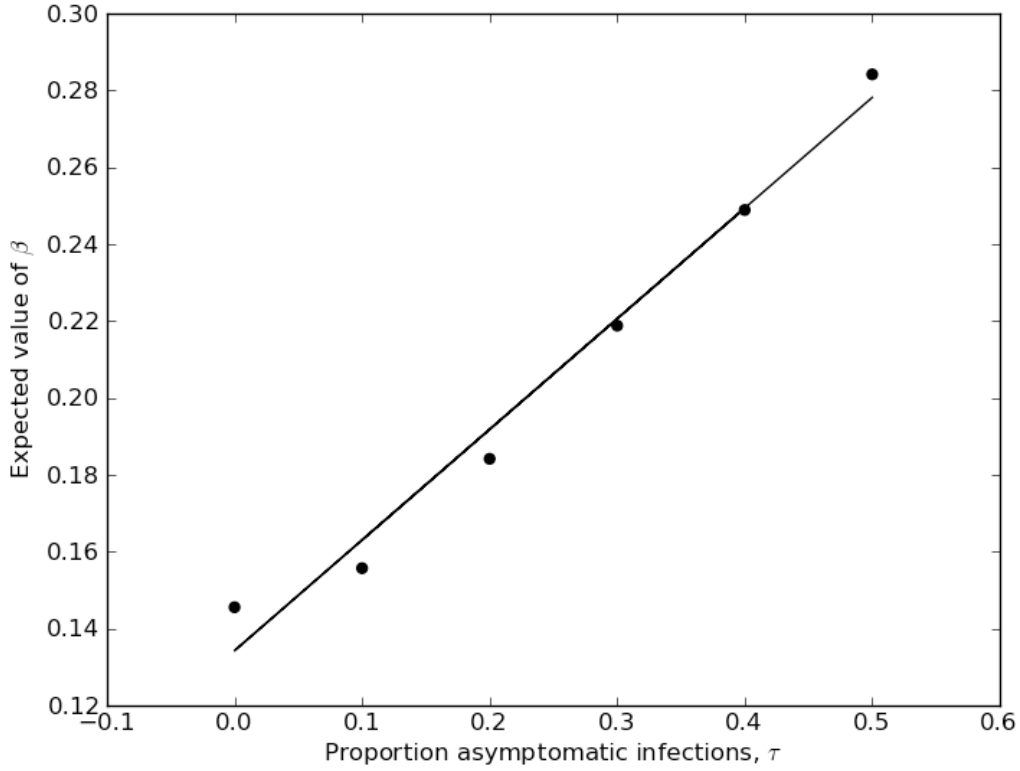
eAlgorithm 2

Asymptomatic infections are, in this simplified model, immediately moved to the immune class. This is because they are significantly less infectious than symptomatic infections, e.g., (10), and can be expected to generate cases on a longer timescale than our window of observation (9 days). Although they are unlikely to contribute significantly to observed within-household transmission dynamics, we expect that they are important to

the community-level persistence of norovirus and, as such, need to be accounted for in the estimate of rate of transmission. In this context, then, asymptomatic cases can be thought of as censored data that bias our estimate of the force of infection.

When simulating outbreaks, we fix the background infection rate and the distribution of the incubation and infectious periods,($\alpha = 0.001$, $1/e = 1.7$ days, $e_S = 4.0$, $1/g = 1.14$ days, $g_S = 1.0$) and allow the transmission parameter, β, and proportion of asymptomatic infections, τ, to vary. We then sample all 126 parameter combinations from $\beta = \{.10, .11, \ldots, .30\}$ and $\tau = \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. We draw 20 stochastic realizations of each parameter set and estimate the mean ML value of β (i.e., average over the 20 runs) for each (τ, β) combination, as though $\tau = 0$. This gives a predicted value of β for each level of τ. Starting from our ML estimate of 0.14 for β when $\tau = 0$, the predicted value of β increases linearly by 0.035 units for each 10% for increase in τ (Figure 8).

We test the sensitivity of these results to the assumption that asymptomatic individuals do not contribute to household transmission by allowing asymptomatic infections to be 10% as infectious as symptomatic ones. We find broadly similar results, with the predicted value of β increasing linearly by 0.028 units for each 10% increase in τ (eFigure 1).



eFigure 1. Relationship of proportion asymptomatic to expected value of β when asymptomatic infections are 10% as infectious as symptomatic infections.

3. Missing Household Sizes

Since all households in our dataset consist of two or more people, the minimum household size, *h*, is 2. We start with the empirical distribution of household sizes from a 1990 census of household sizes in Sweden (see eTable), denoted as *C,* where *C(h)* is the probability of observing a household of size *h* in the total population .

If the minimum possible number of individuals, i.e., the number of infections observed in a household, $h_{\min}$, is less than or equal to 2, the entire empirical distribution is used to sample a household size. If $h_{\min}$ = , the number of cases observed is set as the minimum household size, with values smaller than $h_{\min}$ assigned a density of zero. We

assume that the case data provide no additional information on the distribution of the remaining household sizes, so the remaining sizes on the interval $h_{\min} \le h \le 10$ are assigned a uniform density.

This information is combined with the census data in the top row of eTable for each size to generate a distribution from which we can sample household sizes for h ≥ $h_{\min}$:

$$P(h \mid C, h_{\min}) = \frac{C(h)}{\sum_{h \ge h_{\min}}^{10} C(h)}$$

eEquation

In order to sample random variates from this distribution, we compute the conditional CDF of the household size distribution and draw a random number on the interval (0,1], and select the smallest value of *h* where the CDF is less than equal to the random number.

The second row of eTable shows the probability distribution resulting from this sampling procedure. We find that the expected household size increases slightly from 3.73 to 3.87 individuals, with most of this change accounted for by a decrease in the density of households of size 2 to slightly larger ones.

	# Household Members								
	2	3	4	5	6	7	8	9	10
Census Density	0.325	0.193	0.248	0.108	0.027	0.041	0.024	0.017	0.017
Sampled Density	0.283	0.192	0.265	0.115	0.031	0.047	0.027	0.018	0.019

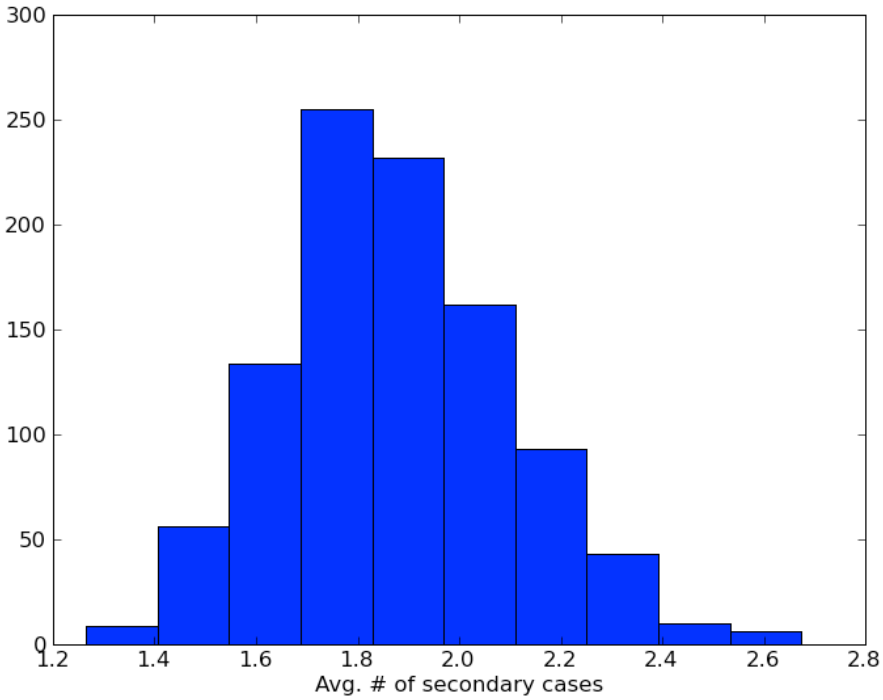
eTable. Empirical Probability Distribution of Household Sizes

4. Model Validation

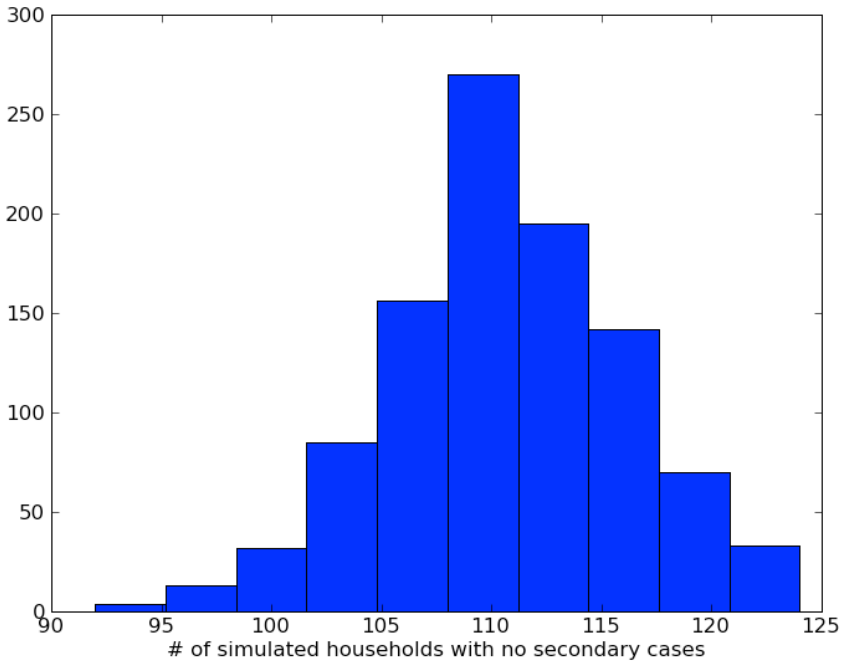
In order to validate the SEIR model used for simulation and parameter estimation, we performed additional simulation analysis using a Gillespie[1] algorithm-based implementation of the model described in eAlgorithm 1, which is an exact, continuous-time simulation of the transmission model.

In each simulation, there are 153 households, the sizes of which are drawn from *C*, the empirical distribution of household sizes. At *t=0*, each household has a single index case. Model parameters are the same as those obtained from our statistical analysis (β = 0.14, $1/\gamma$=1.17 days, γ_s = 1.0). For each of 1000 simulations, we record the number of households with no secondary cases, i.e., where there is stochastic die-out, and the average number of cases in households with secondary cases.

We find that our simulation results are in good agreement with the Stockholm data for both outbreak size (Simulated mean = 1.9 cases, SD = .2, vs. 1.6 for Stockholm data; eFigure 2) and the number of simulated households in which there are no secondary cases (Simulated mean = 110.5 households, SD = 5.5 vs. 104 households for Stockholm data; eFigure 3).



eFigure 2. Histogram of average number of secondary cases in simulated household outbreaks.



eFigure 3. Histogram of number of households with no secondary cases.

5. Computational Details

Data augmentation software was implemented in C++ and Python 2.6 using *Boost.Python* and the *Numpy* and *Scipy* numerical and scientific computing libraries. Plots were generated with *Matplotlib* 0.98 graphing and plotting tools for Python. All diagrams were created in *Inkscape* 0.47.

All results presented here come from 10^4 independent samples for each parameter combination.

References

1.Gillespie, D.T. (1976). "A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions". *Journal of Computational Physics* **22** (4): 403–434