J. Appl. Prob. **42,** 334–345 (2005)
Printed in Israel
© *Applied Probability Trust* 2005

COMPOUND POISSON LIMITS FOR HOUSEHOLD EPIDEMICS

PETER NEAL,* *University of Manchester*

Abstract

We consider epidemics in populations that are partitioned into small groups known as households. Whilst infectious, a typical infective makes *global* and *local* contact with individuals chosen independently and uniformly from the whole population or their own household, as appropriate. Previously, the classical Poisson approximation for the number of survivors of a severe epidemic has been extended to the household model. However, in the current work we exploit a Sellke-type construction of the epidemic process, which enables the derivation of sufficient conditions for the existence of a compound Poisson limit theorem for the survivors of the epidemic. The results are specialised to the *Reed–Frost* and *general stochastic epidemic* models.

Keywords: Epidemic model; household; compound Poisson convergence

2000 Mathematics Subject Classification: Primary 92D30
Secondary 60F05

1. Introduction

Poisson limit theorems for closed population SIR epidemics have a long history dating back to the seminal work [9]. In such epidemic models, individuals are in one of three possible states: susceptible (S), infective (I), or removed (R). The only transitions in state that are assumed to occur are $S \to I$ and $I \to R$, that is a susceptible becoming infected and an infective being removed, respectively. Since [9], an extensive range of asymptotic Poisson limit theorems for homogeneously mixing epidemics have been derived (see [4], [10]–[13]).

However, over the last ten years or so, there has been considerable interest in studying heterogeneously mixing populations (see, for example, [1], [5], [7]). In particular, considerable attention has been focused on two-level mixing models, in which, whilst infectious, individuals make global infectious contact with individuals chosen independently and uniformly from the whole population and local infectious contact with individuals according to some predefined local contact distribution. The most widely studied two-level mixing model is the so-called household model, where individuals belong to disjoint groups known as households, and infectious individuals make local infectious contact with individuals chosen independently and uniformly from their own household.

All three of the 'classic' asymptotic approximations, as the population size $N \to \infty$ and for homogeneously mixing SIR epidemics, have been extended to the households model; that is, the branching process approximation for the early stages of the epidemic (see [3]), a central limit theorem for the final size of the epidemic when the epidemic takes off (see [5]), and a Poisson limit theorem for the total number of individuals who avoid infection during the course of an epidemic that is well above threshold (see [6]). In particular, in [6] it was shown in establishing

Received 10 August 2004; revision received 24 November 2004.
* Postal address: School of Mathematics, University of Manchester, Sackville Street, Manchester M60 1QD, UK.
Email address: p.neal-2@manchester.ac.uk

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

the asymptotic Poisson limit theorem that, in the limit as the population size $N \to \infty$, all the survivors of the epidemic are in distinct households, where a survivor is defined to be an individual who remains susceptible throughout the course of the epidemic. This is because, under the conditions of [6, Theorems 3.1 and 4.6], the global infection process plays a key role in spreading the disease. However, it is reasonable to suggest that, under suitable conditions, the survivors of a severe epidemic will be clustered together in a few households, thus suggesting the possibility of an asymptotic compound Poisson limit theorem. Furthermore, in [6, Section 5], it is stated without proof that, under certain conditions, an asymptotic compound Poisson limit theorem can be derived for the Reed–Frost household model. The aim of this paper is to prove that result. In fact, we prove a far more general asymptotic compound Poisson limit theorem for SIR household epidemics (see Theorem 3.1, below). In order to do this we take a radically different approach to [6], which built upon [10]. Our method is more akin to [13], which gave the first rigorous Poisson limit for the general stochastic epidemic (GSE). In particular, we utilise a Sellke-type construction of the household epidemic (see, for example, [5]). Unlike in [6], we shall assume for ease of exposition that there is only one initial infective. However, the theorems and proofs of Section 3, below, can be easily adapted to the case where there are c initial infectives, for some $1 \le c < \infty$.

In Section 2, a full description of the epidemic model is given. Moreover, two equivalent descriptions of the epidemic model are given, both of which will prove useful in the sequel. In Section 3, the main result, Theorem 3.1, is stated and then proved via a series of lemmas. Finally, in Section 4, we consider two important special cases, the Reed–Frost and GSE models. We compare and contrast the asymptotic compound Poisson limits obtained in these cases.

2. Household epidemic

We consider a sequence of epidemics $\{E_n\}$, indexed by the number of households n. For fixed $n \ge 1$, we shall label the households 1 through to n and, for $1 \le i \le n$, let H_i^n denote the size of household i. We assume that, for all $l \ge 1$ and $n \ge l$, $H_l^n = H_l$, say. For $1 \le i \le n$, the individuals in household i are labelled $(i, 1)$ through to (i, H_i). We assume that there exists an m and an M, $2 \le m \le M$, such that, for all $l \ge 1$, $m \le H_l \le M$. That is, there is a minimum, m, and maximum, M, household size. Whilst the assumption that the minimum household size is at least 2 is slightly restrictive, it is necessary to prevent the compound Poisson limit from being degenerate, in that if $m = 1$ then the only compound Poisson limit that exists is the standard Poisson limit (cf. [6, Section 4.3]). For $n \ge 1$ and $m \le r \le M$, let θ_r^n denote the proportion of households of size r and assume that $\theta_r^n \to \theta_r$ as $n \to \infty$, for some $0 \le \theta_r \le 1$. Finally, for $n \ge 1$, let $N_n = \sum_{i=1}^n H_i^n$ denote the total number of individuals in the population.

The epidemic process can be constructed as follows. Consider a fixed $n \ge 1$. Assign to each individual in the population an independent and identically distributed (i.i.d.) life history according to $\mathcal{H}^n = (Q, \eta^n, \zeta^n)$, where Q is the infectious period, whose distribution is assumed to be independent of n, and η^n and ζ^n are homogeneous Poisson point processes of times, relative to an individual's infection, at which the individual makes global and within-household (local) infectious contact, respectively. Let η^n and ζ^n have rates $\lambda_n^{\mathrm{G}} = N_n \beta_n^{\mathrm{G}}$ and $\lambda_n^{\mathrm{H}} = (m-1)\beta_n^{\mathrm{H}}$, respectively, where each 'β' denotes the useful fraction of the corresponding rate 'λ'. The elements Q, η^n, and ζ^n of $\mathcal{H}^n$ are assumed to be independent. Each global infectious contact is, with an individual, chosen independently and uniformly from the N_n individuals in the population. Each local infectious contact made by an infective (i, j), say, is, with an individual,

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

chosen independently and uniformly from the other $H_i - 1$ individuals in household i. Thus, we assume that individuals make both global and local contact with the other members of their household and that an individual makes global but not local infectious contact with itself. At the end of its infectious period, an individual becomes immune to further infection and plays no role in the remainder of the epidemic. If an infectious individual contacts a susceptible (i, j) with life history $\mathcal{H}^n_{i,j} = (Q_{i,j}, \eta^n_{i,j}, \zeta^n_{i,j})$ at time t, say, then the susceptible becomes infected and is infectious until time $t + Q_{i,j}$, at which time the individual becomes removed. Whilst infectious, individual (i, j) makes global and local infectious contacts at the points $t + \eta^n_{i,j}$ and $t + \zeta^n_{i,j}$, respectively. We assume that there is one initial infective in an otherwise initially susceptible population. For convenience, we assume that the initial infective is individual $(1, 1)$ and that it becomes infected at time $t = 0$. These assumptions on the initial infective can easily be relaxed.

In order to prove Theorem 3.1, below, it is useful to give an alternative description of the epidemic. To this end, we give a Sellke-type construction of the epidemic (see [5] and [13]), which can easily be shown to be equivalent to the above description of the epidemic (see, for example, [5]). Consider fixed $n \geq 1$. Assign to each individual in the population an i.i.d. life history according to $\tilde{\mathcal{H}} = (Q, \Delta^n, \zeta^n)$, where Q and ζ^n are as above and Δ^n is the total exposure to global infection required to infect an individual globally. That is, any individual (i, j), say, is globally infected when exposed to t units of global infectious pressure if and only if $\Delta^n_{i,j} \leq t$, where $\Delta^n_{i,j}$ is distributed according to Δ^n. We take Δ^n to follow a negative exponential distribution with mean $(\beta^{\mathrm{G}}_n)^{-1}$, to be consistent with the previous description of the epidemic. We again assume that the initial infective is individual $(1, 1)$, so $\Delta^n_{1,1}$ is redundant. For notational convenience, we shall take $\Delta^n_{1,1} = 0$. The local infectious process is exactly as described above. Furthermore, we assume that local infections, where they occur, are instantaneous. Whilst this alters the time course of the epidemic, the final outcome of the epidemic is unaffected. This approach has previously been taken in [5].

For $n \geq 1$, $1 \leq i \leq n$, and $t \geq 0$, let $(R^n_i(t), A^n_i(t))$ be defined as follows. Let $R^n_i(t)$ denote the total number of individuals in household i who avoid infection given that the whole population is exposed to t units of global infectious pressure. Let $A^n_i(t)$ denote the sum of the infectious periods of the individuals in household i who are infected when the whole population is exposed to t units of global infectious pressure. Thus, $R^n_i(t)$ and $A^n_i(t)$ do not only take account of those individuals who are globally infected by t units of global infectious pressure but also the subsequent local epidemics in household i originating from those individuals who are infected globally. Note that household 1 is different to the other households in that it contains the initial infective.

We follow [5, Section 4.2.2] in defining a sequence of stochastic times at which to consider the epidemic. Let $I^0_n = Q_{1,1}$ and, for $k \geq 1$, let

$$I^k_n = \sum_{i=1}^{n} A^n_i(I^{k-1}_n).$$

Therefore, I^0_n is just the infectious period of the initial infective and I^1_n is the sum of the infectious periods of all those individuals infected globally by I^0_n units of infectious pressure and also those individuals infected in the subsequent local epidemics. This process can be continued, with I^k_n being the sum of the infectious periods of all those individuals either infected globally or by the subsequent local epidemics when the whole population is exposed to I^{k-1}_n units of global infectious pressure. The process continues until the additional infectious pressure

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

generated by a set of local epidemics is insufficient to infect any further individuals globally. Then $k^* = \min\{k\colon I_n^{k+1} = I_n^k\}$ is well defined since the population is finite. Let $I_n^\infty = I_n^{k^*}$; then I_n^∞ is given by

$$I_n^\infty = \min\left\{t \geq 0\colon t = \sum_{i=1}^n A_i^n(t)\right\}.$$

For $n \geq 1$, let $X_n(t) = \sum_{i=1}^n R_i^n(t)$ and note that $X_n(t)$ is decreasing in t. For $n \geq 1$, let Z_n denote the total number of survivors of the epidemic E_n. Then $Z_n = X_n(I_n^\infty)$. For $n \geq 1$ and $s, t \geq 0$, let $f_n(s;t) = \mathrm{E}[s^{X_n(t)}]$ denote the probability generating function for $X_n(t)$. Note that, for any $t \geq 0$, conditional upon each individual in the population being exposed to t units of global infectious pressure, the local epidemics within each household are independent. Thus,

$$f_n(s;t) = \mathrm{E}[s^{X_n(t)}] = \prod_{i=1}^n \mathrm{E}[s^{R_i^n(t)}]. \tag{2.1}$$

Furthermore, for all $n \geq 2$ and for $2 \leq j, k \leq n$, we obtain $R_j^n(t) \stackrel{\mathrm{D}}{=} R_k^n(t)$ if $H_j = H_k$, where '$\stackrel{\mathrm{D}}{=}$' denotes equality in distribution.

3. Main results

For $\lambda > 0$, let $\mathrm{Po}(\lambda)$ denote a Poisson random variable with mean λ and, for $k \in \mathbb{N}$, let $\mathrm{Po}_k(\lambda) \stackrel{\mathrm{D}}{=} k\mathrm{Po}(\lambda)$.

Theorem 3.1. *Suppose that there exist β, α, and d with $\beta > 0$, $\alpha \in \mathbb{R}$, and $0 < d < \infty$, such that $\mathrm{E}[Q^{1+\beta}] < \infty$,*

$$m\,\mathrm{E}[Q]\lambda_n^{\mathrm{G}} - \log n \to \alpha, \quad \textit{as } n \to \infty, \tag{3.1}$$

and

$$n^{1/m}\varphi_n(1) \to d, \quad \textit{as } n \to \infty, \tag{3.2}$$

where, for $t \geq 0$, $\varphi_n(t) = \mathrm{E}[\exp(-t\beta_n^{\mathrm{H}}Q)]$. Furthermore, for all r with $m \leq r \leq M$ and for all i with $1 \leq i \leq m$, let

$$\hat{d}_{r,m,i} = \lim_{n\to\infty} n^{(m-i)/m}\phi_n\left(\frac{i}{r-1}\right)^{r-i},$$

where, for $t \geq 0$, $\phi_n(t) = \mathrm{E}[\exp(-t\lambda_n^{\mathrm{H}}Q)]$. Then there exist independent random variables $Z^1, Z^2, \ldots, Z^m$ such that

$$Z_n \stackrel{\mathrm{D}}{\to} Z = \sum_{k=1}^m Z^k, \quad \textit{as } n \to \infty,$$

where $Z^k = \mathrm{Po}_k(b_k)$, $b_k = a^k \sum_{r=m}^M \binom{r}{k}\theta_r\hat{d}_{r,m,k}$ $(1 \leq k \leq m)$, $a = \exp(-\alpha/m)$, and '$\stackrel{\mathrm{D}}{\to}$' denotes convergence in distribution.

We postpone proving Theorem 3.1 until the end of this section. We begin by examining the conditions of Theorem 3.1. Firstly, the total number of survivors of the epidemic will be greater than or equal to the total number of households in which every member of the household avoids global infection during the course of the epidemic. Thus, we require that the global infection

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

rate is such that the probability that a household completely avoids infection tends to 0 as $n \to \infty$. This is the key role played by (3.1). By Jensen's inequality, $\varphi_n(1) \geq \exp(-\beta_n^{\mathrm{H}} \mathrm{E}[Q])$. Therefore, (3.2) and $\mathrm{E}[Q] < \infty$ imply that $\beta_n^{\mathrm{H}} \to \infty$ and, hence, $\lambda_n^{\mathrm{H}} \to \infty$ as $n \to \infty$. Thus, to establish a compound Poisson limit for a household epidemic, we require that both $\lambda_n^{\mathrm{G}} \to \infty$ and $\lambda_n^{\mathrm{H}} \to \infty$ as $n \to \infty$.

For $1 \leq i \leq m-1$, let $d_i = \lim_{n\to\infty} n^{1/m}\varphi_n(i)$ and note that $d_i \leq d$. Also, by Jensen's inequality, for $m \leq r \leq M$ and $1 \leq i \leq m-1$,

$$\phi_n\left(\frac{i}{r-1}\right)^{r-i} \leq \phi_n\left(\frac{i}{m-1}\right)^{((m-1)/(r-1))(r-i)} \leq \phi_n\left(\frac{i}{m-1}\right)^{m-i} = \varphi_n(i)^{m-i} \tag{3.3}$$

since, for $t \geq 0$, $\phi_n(t) \leq 1$ and, for $r \geq m$, $(m-i)(r-1) \leq (r-i)(m-1)$. Hence, (3.2) and (3.3) ensure that, for all $m \leq r \leq M$ and $1 \leq i \leq m-1$, $\hat{d}_{r,m,i} \leq d_i^{m-i} < \infty$. Also $\hat{d}_{m,m,m} = 1$ and $\hat{d}_{r,m,m} = 0$ $(m < r \leq M)$. Thus, it follows from (3.1) and (3.2) that $\mathrm{E}[Z] < \infty$.

In order to prove Theorem 3.1, we require the following preliminary lemmas. Throughout we shall assume that the conditions of Theorem 3.1 hold.

Lemma 3.1. *For any* $0 < \varepsilon < 1$,

$$\mathrm{P}(Z_n > \varepsilon N_n) \to 0, \quad \textit{as } n \to \infty.$$

Proof. For $n \geq 1$, let $\tilde{E}_n$ denote the homogeneous mixing epidemic in a population of size N_n, where a typical individual i, say, is infectious for a time Q_i and makes infectious contacts at the points η_i^n. Also, let $\tilde{Z}_n$ denote the total number of survivors of the epidemic. Clearly, $\tilde{Z}_n$ is stochastically greater than $\tilde{Z}_n$, since $\tilde{E}_n$ can be constructed from E_n by simply ignoring the local infectious contacts.

Then, by (3.1), $\lambda_n^{\mathrm{G}} \to \infty$ as $n \to \infty$ and, since $\varphi_n(1) \geq \mathrm{P}(Q = 0)$, by (3.2), $\mathrm{P}(Q = 0) = 0$. Therefore, it can be proved along similar lines to [6, Lemma 3.9] that $\mathrm{P}(\tilde{Z}_n > \varepsilon N_n) \to 0$, as $n \to \infty$, and the lemma follows.

Let $\bar{I}_n = N_n \mathrm{E}[Q]$. For $i \geq 1$, let $\tilde{Q}_i$ be i.i.d. according to Q and, for $K \geq 1$, let $S_K = \sum_{i=1}^K \tilde{Q}_i$.

Lemma 3.2. *For any* $0 < \varepsilon < 1$,

$$\mathrm{P}(I_n^\infty < \varepsilon \bar{I}_n) \to 0, \quad \textit{as } n \to \infty.$$

Proof. Fix $0 < \varepsilon < 1$. Then

$$\mathrm{P}(I_n^\infty < \varepsilon \bar{I}_n) \leq \mathrm{P}\left(I_n^\infty < \varepsilon \bar{I}_n \;\middle|\; Z_n < \frac{1-\varepsilon}{2} N_n\right) + \mathrm{P}\left(Z_n > \frac{1-\varepsilon}{2} N_n\right). \tag{3.4}$$

Lemma 3.1 ensures that the second term on the right-hand side of (3.4) converges to 0 as $n \to \infty$.

Note that $Z_n < \frac{1}{2}(1-\varepsilon)N_n$ implies that the total number of infectives during the course of the epidemic exceeds $\frac{1}{2}(1+\varepsilon)N_n$. Thus, conditional upon $Z_n < \frac{1}{2}(1-\varepsilon)N_n$, I_n^∞ is greater than or equal to the sum of the infectious periods of the first K_n infectives, where $K_n = \lfloor \frac{1}{2}(1+\varepsilon)N_n \rfloor + 1$. (For $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the smallest integer less than or equal to x.) Furthermore, conditional upon $Z_n < \frac{1}{2}(1-\varepsilon)N_n$, we have that, for $1 \leq j \leq K_n$, the first j infectives are responsible for at least j successful infectious contacts. It is then fairly

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

straightforward to adapt the proof of [6, Lemma 3.2] to show that the sum of the infectious periods of the first K_n infectives is stochastically larger than S_{K_n}. Thus,

$$\mathrm{P}\left(I_n^{\infty} < \varepsilon \bar{I}_n \;\middle|\; Z_n < \frac{1-\varepsilon}{2} N_n\right) \le \mathrm{P}(S_{K_n} < \varepsilon \bar{I}_n). \tag{3.5}$$

However, the strong law of large numbers ensures that the right-hand side of (3.5) converges to 0 as $n \to \infty$, since $\mathrm{E}[S_{K_n}] = (\lfloor \frac{1}{2}(1+\varepsilon)N_n \rfloor + 1)\,\mathrm{E}[Q]$. Hence, the lemma is proved.

For $n \ge 1$, let $\pi_n = \exp(-\beta_n^{\mathrm{G}} \bar{I}_n) = \exp(-\lambda_n^{\mathrm{G}}\,\mathrm{E}[Q])$. For $n \ge 1$, $m \le r \le M$, $0 \le j \le r$, and $0 \le k \le r-j$, let $P_{j,k}^{n,r}$ denote the probability that, in a household of size r with j individuals infected globally, there is a local epidemic in which k of the individuals who avoided global infection are infected. Note that, by definition, for all $n \ge 1$ and $m \le r \le M$, $P_{0,0}^{n,r} = 1$. Explicit expressions for the probabilities $\{P_{j,k}^{n,r}\}$ can be obtained using [2, Equation (2.5)].

Lemma 3.3. *Let $\{u_n\}$ be any sequence of positive real numbers such that $u_n \to 1$ as $n \to \infty$. Then, for all $m \le r \le M$ and for all $1 \le j \le m$, we obtain*

$$n(\pi_n u_n)^j P_{r-j,0}^{n,r} \to a^j \hat{d}_{r,m,j}, \quad \text{as } n \to \infty, \tag{3.6}$$

and, for $1 \le k \le j-1$,

$$n(\pi_n u_n)^j P_{r-j,k}^{n,r} \to 0, \quad \text{as } n \to \infty. \tag{3.7}$$

Finally, for all $m < r \le M$, $m < j \le r$, and $0 \le k \le j-1$,

$$n(\pi_n u_n)^j P_{r-j,k}^{n,r} \to 0, \quad \text{as } n \to \infty. \tag{3.8}$$

Proof. By (3.1), for $n \ge 1$, there exists a μ_n such that

$$\lambda_n^{\mathrm{G}} = \frac{1}{m\,\mathrm{E}[Q]}(\log n + \alpha) + \mu_n,$$

where $\mu_n \to 0$ as $n \to \infty$. Therefore, it follows that $\pi_n = an^{-1/m}\exp(-\mu_n\,\mathrm{E}[Q])$, so

$$n^{1/m}\pi_n u_n \to a, \quad \text{as } n \to \infty. \tag{3.9}$$

By [2, Equation (2.5)] and (3.3),

$$n^{(m-j)/m} P_{r-j,0}^{n,r} = n^{(m-j)/m}\phi_n\left(\frac{j}{r-1}\right)^{r-j} \to \hat{d}_{r,m,j}, \quad \text{as } n \to \infty. \tag{3.10}$$

Thus, (3.6) follows from (3.9) and (3.10).

Furthermore, by [2, Equation (2.5)], for $1 \le j \le m$ and $1 \le k \le j-1$,

$$P_{r-j,k}^{n,r} \le \binom{j}{k}\phi_n\left(\frac{j-k}{r-1}\right)^{r-(j-k)}.$$

Thus, by (3.3),

$$P_{r-j,k}^{n,r} \le \binom{j}{k}\varphi_n(j-k)^{m-(j-k)} \le \binom{j}{k}\varphi_n(1)^{m-(j-k)},$$

so, by (3.2),

$$n^{(m-j)/m} P_{r-j,k}^{n,r} \to 0, \quad \text{as } n \to \infty. \tag{3.11}$$

Thus, (3.7) follows from (3.9) and (3.11).

Finally, (3.8) is trivial since, for $j > m$, $n\pi_n^j \to 0$ as $n \to \infty$.

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

Lemma 3.4. *For any $\delta > 0$,*

$$\mathrm{P}(X_n(I_n^\infty) > n^\delta) \to 0, \quad \textit{as } n \to \infty.$$

Proof. Fix an ε such that $1 - \delta < \varepsilon < 1$. Then

$$\begin{aligned}\mathrm{P}(X_n(I_n^\infty) > n^\delta) &\le \mathrm{P}(X_n(I_n^\infty) > n^\delta \mid I_n^\infty \ge \varepsilon \bar{I}_n) + \mathrm{P}(I_n^\infty < \varepsilon \bar{I}_n) \\ &\le \mathrm{P}(X_n(\varepsilon \bar{I}_n) > n^\delta) + \mathrm{P}(I_n^\infty < \varepsilon \bar{I}_n), \end{aligned} \tag{3.12}$$

since $X_n(I_n^\infty) \le X_n(\varepsilon \bar{I}_n)$ if $I_n^\infty \ge \varepsilon \bar{I}_n$. The second term on the right-hand side of (3.12) converges to 0 as $n \to \infty$, by Lemma 3.2.

By Markov's inequality,

$$\mathrm{P}(X_n(\varepsilon \bar{I}_n) > n^\delta) \le n^{-\delta}\, \mathrm{E}[X_n(\varepsilon \bar{I}_n)] = n^{-\delta} \sum_{k=1}^{n} \mathrm{E}[R_k^n(\varepsilon \bar{I}_n)]. \tag{3.13}$$

Consider a household $l \ge 2$, say, and suppose that $H_l = r$. Note that

$$\mathrm{E}[R_l^n(\varepsilon \bar{I}_n)] \le r\, \mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0). \tag{3.14}$$

For $1 \le i \le n$ and $t \ge 0$, let $G_i^n(t)$ denote the total number of individuals in household i who avoid global infection when each individual in the population is exposed to t units of global infectious pressure. Then,

$$\mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0) = \sum_{j=0}^{r} \mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0 \mid G_l^n(\varepsilon \bar{I}_n) = j)\, \mathrm{P}(G_l^n(\varepsilon \bar{I}_n) = j).$$

Therefore, since $\mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0 \mid G_l^n(\varepsilon \bar{I}_n) = 0) = 0$ and $\mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0 \mid G_l^n(\varepsilon \bar{I}_n) = r) = 1$, we have

$$\mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0) = \sum_{j=1}^{r-1} \sum_{k=0}^{j-1} P_{r-j,k}^{n,r}\, \mathrm{P}(G_l^n(\varepsilon \bar{I}_n) = j) + \mathrm{P}(G_l^n(\varepsilon \bar{I}_n) = r).$$

Note that

$$\mathrm{P}(G_l^n(\varepsilon \bar{I}_n) = j) = \binom{r}{j} \exp(-\beta_n^{\mathrm{G}} \varepsilon \bar{I}_n)^j (1 - \exp(-\beta_n^{\mathrm{G}} \varepsilon \bar{I}_n))^{r-j} = \binom{r}{j} \pi_n^{\varepsilon j} (1 - \pi_n^{\varepsilon})^{r-j}.$$

Thus,

$$\begin{aligned} n^\varepsilon\, \mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \ne 0) &= n^\varepsilon \sum_{j=1}^{r-1} \sum_{k=0}^{j-1} \left\{ P_{r-j,k}^{n,r} \binom{r}{j} \pi_n^{\varepsilon j} (1 - \pi_n^{\varepsilon})^{r-j} \right\} + n^\varepsilon \pi_n^{\varepsilon r} \\ &\le \sum_{j=1}^{r-1} \sum_{k=0}^{j-1} \left\{ \binom{r}{j} (n P_{r-j,k}^{n,r} \pi_n^j)^\varepsilon \right\} + (n \pi_n^r)^\varepsilon. \end{aligned}$$

Now it follows from Lemma 3.3 that there exists a constant C_r, $0 \le C_r < \infty$, such that

$$\lim_{n \to \infty} \left\{ \sum_{j=1}^{r-1} \sum_{k=0}^{j-1} \left\{ \binom{r}{j} (n P_{r-j,k}^{n,r} \pi_n^j)^\varepsilon \right\} + (n \pi_n^r)^\varepsilon \right\} = C_r.$$

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

Hence, for all sufficiently large n,

$$\mathrm{P}(R_l^n(\varepsilon \bar{I}_n) \neq 0) \leq (C_r + 1)n^{-\varepsilon}. \tag{3.15}$$

It is straightforward to adapt the above arguments to show that

$$\mathrm{P}(R_1^n(\varepsilon \bar{I}_n) \neq 0) \leq (C_{H_1} + 1)n^{-\varepsilon}. \tag{3.16}$$

Let $C = \max_{m \leq r \leq M}\{C_r\}$. Hence, by (3.13)–(3.16), for all sufficiently large n,

$$\begin{aligned}\mathrm{P}(X_n(\varepsilon \bar{I}_n) > n^\delta) &\leq n^{-\delta}\,\mathrm{E}[X_n(\varepsilon \bar{I}_n)] \\ &\leq n^{-\delta} \sum_{k=1}^{n} H_k(C+1)n^{-\varepsilon} \\ &\leq M(C+1)n^{1-(\delta+\varepsilon)}.\end{aligned}$$

The lemma follows from (3.12), since $\delta + \varepsilon > 1$.

We are now ready to prove our main result.

Proof of Theorem 3.1. Fix a γ, $0 < \gamma < \beta/(1+\beta)$, and set $\delta = 1-\gamma$. Let $T_n^{\mathrm{L}} = N_n - \lfloor n^\delta \rfloor$ and $T_n^{\mathrm{U}} = N_n$. If $X_n(I_n^\infty) \leq n^\delta$ then I_n^∞ is stochastically larger than $S_{T_n^{\mathrm{L}}}$ (cf. the proof of Lemma 3.2). However, clearly I_n^∞ is stochastically smaller than $S_{T_n^{\mathrm{U}}}$. Therefore, let

$$\hat{I}_n^{\mathrm{L}} = N_n\,\mathrm{E}[Q] - (n^\delta\,\mathrm{E}[Q] + n^{1-\gamma}) = N_n\,\mathrm{E}[Q] - n^{1-\gamma}(\mathrm{E}[Q]+1)$$

and

$$\hat{I}_n^{\mathrm{U}} = N_n\,\mathrm{E}[Q] + (n^\delta\,\mathrm{E}[Q] + n^{1-\gamma}) = N_n\,\mathrm{E}[Q] + n^{1-\gamma}(\mathrm{E}[Q]+1).$$

By the Marcinkiewicz–Zygmund generalization of the strong law of large numbers (see [8, p. 122, Theorem 2]), it is straightforward to show that

$$\mathrm{P}(S_{T_n^{\mathrm{L}}} \geq \hat{I}_n^{\mathrm{L}}) \to 1 \quad \text{and} \quad \mathrm{P}(S_{T_n^{\mathrm{U}}} \leq \hat{I}_n^{\mathrm{U}}) \to 1, \qquad \text{as } n \to \infty.$$

Since, by Lemma 3.4, $\mathrm{P}(X_n(I_n^\infty) > n^\delta) \to 0$ as $n \to \infty$, it then follows that

$$\mathrm{P}(I_n^\infty \geq \hat{I}_n^{\mathrm{L}}) \to 1 \quad \text{and} \quad \mathrm{P}(I_n^\infty \leq \hat{I}_n^{\mathrm{U}}) \to 1, \qquad \text{as } n \to \infty.$$

Hence,

$$\mathrm{P}(X_n(I_n^\infty) \leq X_n(\hat{I}_n^{\mathrm{L}})) \to 1 \quad \text{and} \quad \mathrm{P}(X_n(I_n^\infty) \geq X_n(\hat{I}_n^{\mathrm{U}})) \to 1, \qquad \text{as } n \to \infty.$$

Therefore, since $Z_n = X_n(I_n^\infty)$, to prove the theorem it is sufficient to show that both $X_n(\hat{I}_n^{\mathrm{U}})$ and $X_n(\hat{I}_n^{\mathrm{L}})$ converge, in distribution, to Z as $n \to \infty$.

By (2.1), for all $s, t \geq 0$ we have

$$\begin{aligned}f_n(s;t) &= \mathrm{E}[s^{X_n(t)}] \\ &= \prod_{l=1}^{n}\left\{\sum_{k=0}^{H_l} s^k\,\mathrm{P}(R_l^n(t) = k)\right\} \\ &= \prod_{l=1}^{n}\left\{1 + \sum_{k=1}^{H_l}(s^k - 1)\,\mathrm{P}(R_l^n(t) = k)\right\},\end{aligned} \tag{3.17}$$

since $\mathrm{P}(R_l^n(t) = 0) = 1 - \sum_{k=1}^{H_l} \mathrm{P}(R_l^n(t) = k)$.

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

Consider a household $l \geq 2$, say, with $H_l = r$. Then, for all $1 \leq k \leq r$,

$$\begin{aligned} n\,\mathrm{P}(R_l^n(\hat{I}_n^{\mathrm{U}}) = k) &= n \sum_{j=0}^{r} \mathrm{P}(R_l^n(\hat{I}_n^{\mathrm{U}}) = k \mid G_l^n(\hat{I}_n^{\mathrm{U}}) = j)\,\mathrm{P}(G_l^n(\hat{I}_n^{\mathrm{U}}) = j) \\ &= n \sum_{j=k}^{r} P_{r-j,j-k}^{n,r}\,\mathrm{P}(G_l^n(\hat{I}_n^{\mathrm{U}}) = j). \end{aligned}$$

For $n \geq 1$, let $u_n = \exp(-\beta_n^{\mathrm{G}} n^{1-\gamma}(1 + \mathrm{E}[Q]))$. Then $\exp(-\beta_n^{\mathrm{G}} \hat{I}_n^{\mathrm{U}}) = u_n \pi_n$, where $u_n \to 1$ as $n \to \infty$. Thus, by Lemma 3.3, it follows that, for $1 \leq k \leq m$,

$$\begin{aligned} n\,\mathrm{P}(R_l^n(\hat{I}_n^{\mathrm{U}}) = k) &= n \sum_{j=k}^{r} P_{r-j,j-k}^{n,r} \left\{ \binom{r}{j} (u_n \pi_n)^j (1 - u_n \pi_n)^{r-j} \right\} \\ &\to \binom{r}{k} a^k \hat{d}_{r,m,k}, \quad \text{as } n \to \infty, \end{aligned} \tag{3.18}$$

and, for $m < k \leq H_l$,

$$n\,\mathrm{P}(R_l^n(\hat{I}_n^{\mathrm{U}}) = k) \to 0, \quad \text{as } n \to \infty. \tag{3.19}$$

For all $n \geq 1$, let $\tilde{\theta}_{H_1}^n = \theta_{H_1}^n - 1/n$ and $\tilde{\theta}_r^n = \theta_r^n$ for $r \neq H_1$. It is trivial to show that

$$\mathrm{P}(R_1^n(\hat{I}_n^{\mathrm{U}}) \neq 0) \to 0, \quad \text{as } n \to \infty.$$

Hence, for all $s \geq 0$,

$$\mathrm{E}[s^{R_1^n(\hat{I}_n^{\mathrm{U}})}] \to 1, \quad \text{as } n \to \infty. \tag{3.20}$$

Therefore, it follows from (3.17)–(3.20) that

$$\begin{aligned} f_n(s; \hat{I}_n^{\mathrm{U}}) &= \prod_{l=1}^{n} \left\{ 1 + \sum_{k=1}^{H_l} (s^k - 1)\,\mathrm{P}(R_l^n(\hat{I}_n^{\mathrm{U}}) = k) \right\} \\ &= \mathrm{E}[s^{R_1^n(\hat{I}_n^{\mathrm{U}})}] \\ &\quad \times \prod_{r=m}^{M} \left\{ 1 + \sum_{k=1}^{r} (s^k - 1) \left(\sum_{j=k}^{r} P_{r-j,j-k}^{n,r} \left\{ \binom{r}{j} (u_n \pi_n)^j (1 - u_n \pi_n)^{r-j} \right\} \right) \right\}^{\tilde{\theta}_r^n n} \\ &\to \prod_{r=m}^{M} \exp\left(\theta_r \sum_{k=1}^{m} (s^k - 1) \binom{r}{k} a^k \hat{d}_{r,m,k} \right) \\ &= \prod_{k=1}^{m} \exp\left((s^k - 1) a^k \sum_{r=m}^{M} \theta_r \binom{r}{k} \hat{d}_{r,m,k} \right) \\ &= \mathrm{E}[s^Z]. \end{aligned}$$

Hence, $X_n(\hat{I}_n^{\mathrm{U}}) \xrightarrow{\mathrm{D}} Z$ as $n \to \infty$.

Since $\exp(-\beta_n^{\mathrm{G}} \hat{I}_n^{\mathrm{L}}) = \pi_n u_n^{-1}$, we can similarly show that $X_n(\hat{I}_n^{\mathrm{L}}) \xrightarrow{\mathrm{D}} Z$ as $n \to \infty$. Hence, the theorem is proved.

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

Suppose that all households are of equal size m, say. Then $Z = \sum_{k=1}^{m} Z^k = \sum_{k=1}^{m} \mathrm{Po}_k(b_k)$ with $b_k = \binom{m}{k} a^k d_k^{m-k}$, $1 \le k \le m$, and $d_i = \lim_{n\to\infty} n^{1/m}\varphi_n(i)$, $1 \le i \le m$.

Note that, as well as being able to derive a compound Poisson limit theorem for the total number of survivors of the epidemic, we can use the approach taken in proving Theorem 3.1 to prove other results concerning the final outcome of the epidemic. In particular, in Theorem 3.2, we consider the distribution of the number of households of a given size with a given number of survivors. Therefore, for $n \ge 1$, $m \le r \le M$, and $1 \le u \le r$, let $Y_{r,u}^n$ denote the number of households of size r containing u survivors. Note that, as shown in the proof of Theorem 3.1, in the limit as $n \to \infty$, there are no households with more than m survivors. Thus, for $r > m$ and $m < u \le r$, $Y_{r,u}^n \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, where '$\xrightarrow{\mathrm{P}}$' denotes convergence in probability. Let $\boldsymbol{Y}_r^n = (Y_{r,1}^n, Y_{r,2}^n, \dots, Y_{r,m}^n)$ and let $\boldsymbol{Y}^n = (\boldsymbol{Y}_m^n, \boldsymbol{Y}_{m+1}^n, \dots, \boldsymbol{Y}_M^n)$.

Theorem 3.2. *For r and u such that $m \le r \le M$ and $1 \le u \le m$, let $Y_{r,u}$ be independently distributed Poisson random variables with $Y_{r,u} \sim \mathrm{Po}(\binom{r}{u}\theta_r a^u \hat{d}_{r,m,u})$ and let $\boldsymbol{Y}_r = (Y_{r,1}, Y_{r,2}, \dots, Y_{r,m})$ with $\boldsymbol{Y} = (\boldsymbol{Y}_m, \boldsymbol{Y}_{m+1}, \dots, \boldsymbol{Y}_M)$. Then, under the conditions of Theorem 3.1, we obtain*

$$\boldsymbol{Y}^n \xrightarrow{\mathrm{D}} \boldsymbol{Y}, \quad \text{as } n \to \infty.$$

Proof. The proof of this theorem is similar to that of Theorem 3.1; hence, the details are omitted.

4. Special cases

The conditions of Theorem 3.1, as with the conditions of [6, Theorem 4.6], are simple and easy to check. However, depending upon the distribution of Q, very different compound Poisson limits can be derived. To demonstrate this, we consider probably the two most studied epidemic models, the Reed–Frost and GSE models.

For the Reed–Frost model, it is assumed that all infectious periods are of constant length, that is, $Q \equiv q > 0$. Thus, the following corollary gives a proof of a generalization of the result stated in [6, Section 5].

Corollary 4.1. *Suppose that $Q \equiv q$ for some $q > 0$. Suppose that there exist α and d, with $\alpha \in \mathbb{R}$ and $0 < d < \infty$, such that*

$$mq\lambda_n^{\mathrm{G}} - \log n \to \alpha, \quad \text{as } n \to \infty,$$

and

$$mq\lambda_n^{\mathrm{H}} - (m-1)\log n \to -m(m-1)\log d, \quad \text{as } n \to \infty. \tag{4.1}$$

Then, as $n \to \infty$,

$$Z_n \xrightarrow{\mathrm{D}} \mathrm{Po}(b_1) + \mathrm{Po}_m(b_m),$$

where $b_1 = ad^{m-1}\sum_{r=m}^{M}\theta_r r$ and $b_m = \theta_m a^m$.

Proof. Since $\lambda_n^{\mathrm{H}} = (m-1)\beta_n^{\mathrm{H}}$ and, for $t \ge 0$, $\varphi_n(t) = \exp(-t\beta_n^{\mathrm{H}} q)$, it is trivial to rearrange (4.1) to show that

$$n^{1/m}\varphi_n(1) \to d, \quad \text{as } n \to \infty.$$

Note that, for all r and k such that $m \le r \le M$ and $1 \le k \le m-1$,

$$\phi_n\left(\frac{k}{r-1}\right) = \varphi_n(k)^{(m-1)/(r-1)}.$$

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

Also, for $k \geq 1$,

$$\lim_{n\to\infty} n^{k/m}\varphi_n(k) = \lim_{n\to\infty} n^{k/m}\varphi_n(1)^k = d^k.$$

Therefore, for all $m \leq r \leq M$, $\hat{d}_{r,m,1} = d^{m-1}$ and $\hat{d}_{r,m,k} = 0$ $(2 \leq k \leq m-1)$. The corollary then follows from Theorem 3.1.

For the GSE model, $Q \sim \mathrm{Exp}(\gamma)$, for some $\gamma > 0$.

Corollary 4.2. *Suppose that $Q \sim \mathrm{Exp}(\gamma)$ for some $\gamma > 0$. Suppose that there exist α and d, with $\alpha \in \mathbb{R}$ and $0 < d < \infty$, such that*

$$m\lambda_n^{\mathrm{G}} - \gamma \log n \to \gamma\alpha, \quad \text{as } n \to \infty, \tag{4.2}$$

and

$$d\lambda_n^{\mathrm{H}} n^{-1/m} \to (m-1)\gamma, \quad \text{as } n \to \infty. \tag{4.3}$$

Then, as $n \to \infty$,

$$Z_n \xrightarrow{\mathrm{D}} \sum_{k=1}^{m} Z^k = \sum_{k=1}^{m} \mathrm{Po}_k(b_k),$$

where $b_k = \theta_m \binom{m}{k} a^k (d/k)^{m-k}$, $1 \leq k \leq m$.

Proof. For $t \geq 0$, $\varphi_n(t) = \gamma/(\gamma + t\beta_n^{\mathrm{H}})$. It is therefore trivial to show that (4.3) implies that

$$n^{1/m}\varphi_n(1) \to d, \quad \text{as } n \to \infty.$$

Also, for $m \leq r \leq M$ and $k \geq 1$,

$$\begin{aligned}
\hat{d}_{r,m,k} &= \lim_{n\to\infty} n^{(m-k)/m}\phi_n\left(\frac{k}{r-1}\right)^{r-k} \\
&= \lim_{n\to\infty} n^{(m-k)/m}\left(\frac{\gamma(r-1)}{\gamma(r-1)+k\lambda_n^{\mathrm{H}}}\right)^{r-k} \\
&= \begin{cases} \left(\dfrac{d}{k}\right)^{m-k} & \text{if } r = m, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Since $\mathrm{E}[Q] = 1/\gamma$, (4.2) is equivalent to (3.1), and the corollary follows from Theorem 3.1.

Thus, we observe from Corollaries 4.1 and 4.2 that we get very different compound Poisson limits, depending upon Q. In particular, for the Reed–Frost model, in the limit as $n \to \infty$, the survivors of the epidemic either belong to a household of size m that completely escapes infection or are each the only survivor within their respective households. By contrast, for the GSE model, we can find any number of survivors, from 1 through to m, within a household in the limit as $n \to \infty$. However, in the limit as $n \to \infty$, all the survivors of the epidemic belong to households of size m.

Acknowledgements

I would like to thank Daryl Daley for suggesting the possibility of a compound Poisson limit for household epidemics and Frank Ball for constructive comments on an earlier draft of the paper. Also I would like to thank Ron Doney and an anonymous referee for bringing to my attention [8] and [11], respectively.

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press

References

[1] Andersson, H. (1999). Epidemic models and social networks. *Math. Scientist* **24,** 128–147.

[2] Ball, F. G. (1986). A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Adv. Appl. Prob.* **18,** 289–310.

[3] Ball, F. G. (1996). Threshold behaviour in stochastic epidemics among households. In *Athens Conference on Applied Probability and Time Series* (Lecture Notes Statist. **114**), Vol. 1, *Applied Probability*, eds C. C. Heyde, Y. V. Prohorov, R. Pyke and S. T. Rachev, Springer, New York, pp. 253–266.

[4] Ball, F. G. and Barbour, A. D. (1990). Poisson approximation for some epidemic models. *J. Appl. Prob.* **27,** 479–490.

[5] Ball, F. G., Mollison, D. and Scalia-Tomba, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7,** 46–89.

[6] Ball, F. G. and Neal, P. J. (2004). Poisson approximations for epidemics with two levels of mixing. *Ann. Prob.* **32,** 1168–1200.

[7] Becker, N. G. and Dietz, K. (1995). The effect of the household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127,** 207–219.

[8] Chow, Y. S. and Teicher, H. (1978). *Probability Theory*. Springer, New York.

[9] Daniels, H. E. (1967). The distribution of the total size of an epidemic. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* **4,** 281–293.

[10] Lefèvre, C. and Utev, S. (1995). Poisson approximation for the final state of a generalised epidemic process. *Ann. Prob.* **23,** 1139–1162.

[11] Lefèvre, C. and Utev, S. (1996). Asymptotic behaviour of the final state of the generalised epidemic process. *Siberian Math. J.* **37,** 753–763.

[12] Lefèvre, C. and Utev, S. (1997). Mixed Poisson approximation in the collective epidemic model. *Stoch. Process. Appl.* **69,** 217–246.

[13] Sellke, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *J. Appl. Prob.* **20,** 390–394.

https://doi.org/10.1239/jap/1118777174 Published online by Cambridge University Press