Epidemic size and probability in populations with heterogeneous infectivity and susceptibility

Joel C. Miller[*]

*Mathematical Modeling & Analysis Group and Center for Nonlinear Studies, MS B284, Los Alamos National Laboratory,
Los Alamos, New Mexico, USA*

We analytically address disease outbreaks in large, random networks with heterogeneous infectivity and susceptibility. The transmissibility T_{uv} (the probability that infection of u causes infection of v) depends on the infectivity of u and the susceptibility of v. Initially, a single node is infected, following which a large-scale epidemic may or may not occur. We use a generating function approach to study how heterogeneity affects the probability that an epidemic occurs and, if one occurs, its attack rate (the fraction infected). For fixed average transmissibility, we find upper and lower bounds on these. An epidemic is most likely if infectivity is homogeneous and least likely if the variance of infectivity is maximized. Similarly, the attack rate is largest if susceptibility is homogeneous and smallest if the variance is maximized. We further show that heterogeneity in the infectious period is important, contrary to assumptions of previous studies. We confirm our theoretical predictions by simulation. Our results have implications for control strategy design and identification of populations at higher risk from an epidemic.

The spread of infectious disease is a problem of great interest. Much work has focused on how diseases spread in networks of human, animal, or computer interactions [1–9]. The *transmissibility*, the probability that an edge transmits infection, has a network-dependent threshold, above which epidemics may occur and below which epidemics are not found. Ideally, interventions reduce the transmissibility or modify the network so that epidemics cannot occur. Most study has focused on determining the threshold value under varying assumptions [6–11], in order to design interventions.

For many diseases and networks, it is impractical to eliminate the possibility of an epidemic. An intervention strategy must therefore optimize competing goals: minimize social cost, reduce the probability of an epidemic denoted $\mathcal{P}$, and reduce the attack rate (fraction infected) denoted $\mathcal{A}$ if an epidemic does occur. Most researchers consider just $\mathcal{A}$, but recent investigations [2–4,12,13] have also considered $\mathcal{P}$. No studies systematically investigate the effect of heterogeneity on $\mathcal{P}$ or $\mathcal{A}$, although they can result from variations in the application of interventions, or from natural effects such as variation in recovery time. It is often assumed that the special case of heterogeneous recovery time can be mapped without loss of generality to recovery of all individuals after a single time step [4,6–9,14]. However, it may be inferred from [15] that this assumption is false. We have recently become aware of independent work [16], using techniques similar to, but distinct from, our own, to show that recovery time heterogeneity reduces $\mathcal{P}$ but has no effect on $\mathcal{A}$. In this study, we consider how generic heterogeneities affect $\mathcal{P}$ and $\mathcal{A}$, deriving sharp upper and lower bounds.

The epidemics we study spread on random networks of N nodes with degree k distributed according to $P(k)$. We modify the susceptible-infected-recovered model [1,17] to include heterogeneities: nodes are classified as susceptible, infectious, or recovered. The outbreak begins with a single

infection (the *index case*) which spreads to adjacent nodes. An edge between an infectious node u with infectivity $\mathcal{I}_u$ and a susceptible node v with susceptibility $\mathcal{S}_v$ has transmissibility $T_{uv}=T(\mathcal{I}_u,\mathcal{S}_v)$, and so v is infected with probability T_{uv}. Infectious nodes recover and are no longer susceptible. If an epidemic occurs, the eventual number infected is $O(N)$, otherwise the outbreak is localized. $\mathcal{I}$ and $\mathcal{S}$ can be arbitrary, e.g., $\mathcal{I}$ may be a vector representing duration of infection, level of virus shedding, frequency of handwashing, etc. The form of T is also general: it need only be integrable and bounded in [0,1].

The spread of an epidemic on a network with heterogeneous infectivity and susceptibility is equivalent to a special case of directed percolation for which the probability of retaining an edge depends on both the base and target nodes. In this formalism, infection spreads to the out-component of the index case [2,3]. If the disease has sufficiently high average transmissibility, a single giant strongly connected component G_{scc} exists [18], occupying a fixed fraction of the network as $N \rightarrow \infty$. The set of nodes not in G_{scc}, but from which G_{scc} can be reached, is denoted G_i, while the set of nodes not in G_{scc}, but reachable from G_{scc}, is denoted G_o (i for "in" and o for "out") as shown in Fig. 1. If the index case is in $G_i \cup G_{scc}$, an epidemic occurs, infecting all of $G_o \cup G_{scc}$ and very few other nodes. In the limit $N \rightarrow \infty$, the probability that the index case is in $G_i \cup G_{scc}$, is the probability of an epidemic, and the
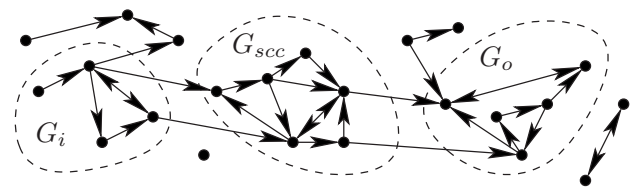

FIG. 1. Schematic representation of G_i, G_{scc}, and G_o. All nodes in G_{scc} can reach any other node in G_{scc}.

*jomiller@lanl.gov

fraction of nodes in $G_o \cup G_{scc}$ is the attack rate.[1] We formally define $\mathcal{P} = \lim_{N\to\infty} |G_i \cup G_{scc}|/N$ and $\mathcal{A} = \lim_{N\to\infty} |G_o \cup G_{scc}|/N$. In general, the sizes of G_i and G_o may differ, so $\mathcal{P} \neq \mathcal{A}$. This contrasts with homogeneous transmissibility, which can be mapped to undirected bond percolation [4,19,20], so $\mathcal{P} = \mathcal{A}$.

We develop a generating function [21] approach to find $\mathcal{P}$, allowing both $\mathcal{I}$ and $\mathcal{S}$ to be heterogeneous. Generating functions have been used to study disease spread in the body [22] or in society [3,4]. Our approach is most similar to that of [22]. We calculate $\mathcal{P}$ based on the distribution of $\mathcal{I}$ and $\mathcal{S}$. Then, holding the average transmissibility fixed, we use Jensen's inequality to find strict upper and lower bounds. G_o and G_i interchange roles under edge reversal, so $\mathcal{A}$ is calculated similarly. We confirm our predictions by simulation.

The infectivity $\mathcal{I}_u$ and susceptibility $\mathcal{S}_u$ of node u are chosen from independent distributions $P(\mathcal{I})$ and $P(\mathcal{S})$. Given $\mathcal{I}_u$, the relation $T(\mathcal{I}_u, \mathcal{S}_v)$, and the distribution $P(\mathcal{S})$, we define the *out-transmissibility* of u as

$$T_o(u) = \int T(\mathcal{I}_u, \mathcal{S}_v) P(\mathcal{S}_v) d\mathcal{S}_v. \tag{1}$$

$P(\mathcal{I})$ and (1) determine the distribution $P_o(T_o)$. We similarly define the *in-transmissibility T_i* and its distribution $P_i(T_i)$. P_o and P_i must have the same average T^* (some pairs P_o and P_i with the same average are inconsistent, but for each P_o, a consistent P_i exists, and vice versa). Henceforth, we use P_i (T_i) and $P_o(T_o)$ rather than $P(\mathcal{S})$ and $P(\mathcal{I})$.

We choose the index case u_0 uniformly from the population. We classify an infected case by its generation, the length of the shortest chain of infectious contacts between it and u_0 (generation 0). Generations may overlap in time, changing temporal dynamics, but leaving $\mathcal{P}$ and $\mathcal{A}$ unchanged.

Our class of random networks is defined by the Molloy-Reed algorithm [23]. Short cycles are rare. The neighborhood of u_0 is treelike on successively longer length scales as $N \to \infty$. Consequently, $\mathcal{P}$ equals the probability that the transmission chains in an infinite tree are infinite.

We define a probability generating function $f(x)$ for the number of infected nodes in generation 1:

$$f(x) = p_0 + p_1 x + \cdots + p_j x^j + \cdots,$$

where p_j is the probability that the index case directly infects j neighbors. The index case has degree k with probability $P(k)$, and thus p_j is given by

$$p_j = \sum_{k=j}^{\infty} P(k) \int_0^1 \mathrm{Bi}(k,j,T_o) P_o(T_o) dT_o,$$

where $\mathrm{Bi}(k,j,T_o)$ is the likelihood of j successful trials from k attempts, each with probability T_o. Note that p_j depends on the distribution P_o but not P_i.

In subsequent generations, the probability of infection is

proportional to degree. Early in the outbreak, an infected node with degree k has $k-1$ susceptible neighbors (the source of its infection cannot be reinfected). The probability that this individual infects j neighbors is

$$q_j = \frac{1}{\langle k \rangle} \sum_{k=j+1}^{\infty} k P(k) \int_0^1 \mathrm{Bi}(k-1,j,T_o) P_o(T_o) dT_o,$$

where $\langle \cdot \rangle$ denotes the expected value. We let $h(x) = \Sigma q_j x^j$ be the generating function for the number of new cases caused by a nonindex case. The generating function for the number of infections in generation $g > 0$ is

$$f(h^{g-1}(x)),$$

where h^{g-1} denotes composition of h with itself $g-1$ times. For later use, we rearrange f and h as

$$f(x) = \int_0^1 P_o(T_o) \sum_{k=0}^{\infty} [1 + T_o(x-1)]^k P(k) dT_o, \tag{2}$$

$$h(x) = \int_0^1 \frac{P_o(T_o)}{\langle k \rangle} \sum_{k=1}^{\infty} [1 + T_o(x-1)]^{k-1} k P(k) dT_o. \tag{3}$$

The extinction probability is $\lim_{g\to\infty} f(h^{g-1}(0))$. To calculate this, we find $\lim_{g\to\infty} h^{g-1}(0)$, which is a solution to $x = h(x)$. At most two solutions exist in the interval $[0,1]$, one of which is $x=1$. If no other solution exists, then $x=1$ is a stable fixed point, and $\mathcal{P}=0$. Otherwise, the iteration converges to $x_0 < 1$ and

$$\mathcal{P} = 1 - f(x_0).$$

Because f and h are independent of P_i, $\mathcal{P}$ is unaffected by heterogeneities in susceptibility.

We now seek distributions P_o to maximize or minimize $\mathcal{P}$ subject to $\langle T \rangle = T^*$. In investigating heterogeneous recovery times [16], showed that identical recovery times maximize $\mathcal{P}$. We use a similar proof to generalize this to arbitrary heterogeneities. For notational convenience, we use $\delta^*(T)$ to denote $\delta(T - T^*)$, set

$$\hat{h}(T,x) = \frac{1}{\langle k \rangle} \sum_{k=1}^{\infty} [1 + T(x-1)]^{k-1} k P(k),$$

and rewrite (3) to explicitly show that h depends on P_o,

$$h[P_o](x) = \int_0^1 \hat{h}(T_o, x) P_o(T_o) dT_o.$$

We similarly define $f[P_o](x)$. Because $\hat{h}$ is a convex function of T, Jensen's inequality shows $P_o = \delta^*$ minimizes $h[P_o](x)$. We denote the smallest root of $x = h[\delta^*](x)$ by x_1. For $x < x_1$ and any P_o, we have $x < h[\delta^*](x) \leq h[P_o](x)$. Thus the root x_0 of $x = h[P_o](x)$ satisfies $x_1 \leq x_0$, so x_0 is minimized if $P_o = \delta^*$.

Similar calculations show that $f[\delta^*](x) \leq f[P_o](x)$ for all P_o. Further, $f[\delta^*](x)$ is an increasing function of x. Thus the extinction probability $f[P_o](x_0)$ is minimized by $P_o = \delta^*$. So homogeneous infectivity maximizes $\mathcal{P}$.

In addition, we find a new lower bound. Jensen's

---

[1] This statement is true for the probability at all N, but if the index case is in G_i, nodes outside $G_o \cup G_{scc}$ are also infected, creating a $o(1)/N$ change to the attack rate.
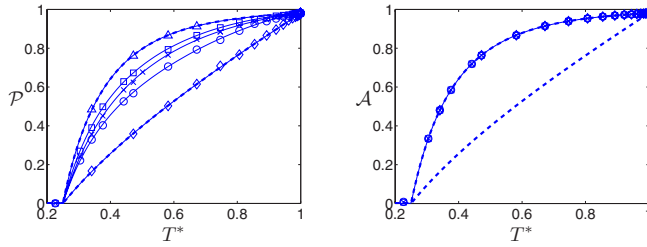
FIG. 2. (Color online) Comparison of theory (curves) with simulation (symbols). For the different distributions of infectivity (with susceptibility constant), $\mathcal{P}$ changes, but $\mathcal{A}$ does not. The theoretical bounds are in dashed bold. We use constant recovery time $\tau=5$ ($\triangle$), $\tau=0$ or ∞ ($\diamond$), $\tau=2$ or 8 ($\square$), $\tau=1$ or 10 ($\bigcirc$), and finally exponentially distributed recovery time ($\times$).

inequality also implies that fixing $\langle T \rangle = T^*$ but increasing $\langle T^2 \rangle$ reduces $\mathcal{P}$. Consequently, $\mathcal{P}$ is minimized by $P_o(T_o) = (1-T^*)\delta(T_o) + T^*\delta(T_o-1)$.

Thus we have shown that an epidemic is most likely if T_o is homogeneous, and least likely if its variance is maximized. Analogously, the attack rate is largest if T_i is homogeneous, and smallest if its variance is maximized. It may be shown that the upper bounds correspond to the relative size of the giant component under bond percolation, while the lower bounds correspond to its relative size under site percolation.

We now find the threshold value of $\langle T \rangle$ above which epidemics can occur [$x=h(x)$ has two roots], and below which they cannot. If we vary $\langle T \rangle$ by continuously changing P_o, the fixed point $x=1$ of $x=h(x)$ bifurcates into two when $h'(1) = 1$. We find

$$h'(1) = R_0 \equiv \langle T \rangle \langle k^2 - k \rangle / \langle k \rangle.$$

So the epidemic threshold is $\langle T \rangle = \langle k \rangle / \langle k^2 - k \rangle$, generalizing [3,4]. R_0 is frequently used in epidemiology [1], representing the average number of new infections an infection causes early in an outbreak.

We confirm our predictions by comparison with simulations on 100 000 node networks. We take an epidemic to occur if over 500 nodes are infected. Our first comparison investigates varying recovery time in an Erdős-Rényi network with $\langle k \rangle = 4$. We discretize time and take different models of recovery time given in the caption of Fig. 2. For each time step, the probability of infecting a susceptible neighbor is p, so an individual with recovery time τ has $T_o = 1 - (1-p)^\tau$. As a reference we take the case where recovery occurs after exactly five time steps. We vary p to change the average transmissibility T^*. The fraction of nodes with each recovery time is chosen so that $\sum P(\tau)[1-(1-p)^\tau] = 1 - (1-p)^5 = T^*$.

We show results in Fig. 2, with 10 000 simulations per symbol. Theory and simulation agree well. The upper bound for epidemic probability occurs when all infections last exactly five time steps ($\triangle$). The lower bound occurs when some nodes remain infectious forever, infecting all neighbors, while the rest infect no one ($\diamond$). Susceptibility is homogeneous, so $\mathcal{A}$ is at the upper bound in all cases.
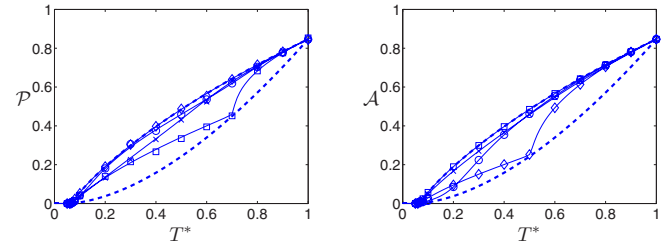

FIG. 3. (Color online) Comparison of theory (curves) with simulation (symbols) for $T_{uv} = 1 - \exp(-\alpha \mathcal{I}_u \mathcal{S}_v)$. The theoretical bounds are in dashed bold. The distributions are $\diamond$, $P(\mathcal{I}) = \delta(\mathcal{I}-1)$, $P(\mathcal{S}) = 0.5\delta(\mathcal{S}-0.001) + 0.5\delta(\mathcal{S}-1)$; $\times$, $P(\mathcal{I}) = 0.5\delta(\mathcal{I}-0.1) + 0.5\delta(\mathcal{I}-1)$, $P(\mathcal{S}) = 0.2\delta(\mathcal{S}-0.1) + 0.8\delta(\mathcal{S}-1)$; $\bigcirc$, $P(\mathcal{I}) = 0.5\delta(\mathcal{I}-0.1) + 0.5\delta(\mathcal{I}-1)$, $P(\mathcal{S}) = 0.8\delta(\mathcal{S}-0.01) + 0.2\delta(\mathcal{S}-1)$; $\square$, $P(\mathcal{I}) = 0.3\delta(\mathcal{I}-0.001) + 0.7\delta(\mathcal{I}-1)$, $P(\mathcal{S}) = \delta(\mathcal{S}-1)$.

Our second comparison has both $\mathcal{I}$ and $\mathcal{S}$ heterogeneous. Following [4], we use a scale-free network with exponential cutoff:[2] $P(k) \propto k^{-2} e^{-k/50}$. The giant component occupies about 85% of the network. We take $T_{uv} = 1 - \exp(-\alpha \mathcal{I}_u \mathcal{S}_v)$ with fixed distributions of scalar $\mathcal{I}$ and $\mathcal{S}$ but varying α to tune T^*. Figure 3 shows theory and simulation in agreement. The upper bounds are achieved for this particular form of T_{uv}. However, by analogy with site percolation, the lower bound of $\mathcal{P}$ ($\mathcal{A}$) occurs only if $T_o=1$ ($T_i=1$) for some nodes, while all others have $T_o=0$ ($T_i=0$). This is almost reached by the $\square$ ($\diamond$) distribution at $T^*=0.7$ ($T^*=0.5$) (cf., the *reinfection threshold* of [24]).

We note that scale-free networks are questionable models of networks on which most diseases spread, because true scale-free networks have $R_0=\infty$, while even emergent diseases have R_0 of only 2–3 [25,26]. Some networks (e.g., sexual networks [27]) may be scale-free but have high assortativity, which is not in this model. Assortativity plays an important role in explaining the anomalous early U.S. HIV epidemic growth [28]. Consequently, the model developed here cannot adequately predict epidemic properties for these networks.

We have shown that a wide class of heterogeneities in infectivity and susceptibility can be studied with generating functions to calculate the epidemic probability $\mathcal{P}$ and attack rate $\mathcal{A}$. We find that $\mathcal{P}$ and $\mathcal{A}$ may differ substantially. In particular, heterogeneity in recovery time has a significant effect on $\mathcal{P}$ and cannot be ignored.

For fixed average transmissibility we have found upper and lower bounds for both $\mathcal{P}$ and $\mathcal{A}$. Further, we have found distributions realizing these bounds. For fixed average transmissibility, increasing the variance of T_o decreases $\mathcal{P}$, and increasing the variance of T_i decreases $\mathcal{A}$.

These results can be used to assist in designing control strategies. For example, if choosing between a strategy that reduces infectivity or susceptibility by half for all of the population or one that reduces infectivity or susceptibility completely for half the population, it is better to choose the latter. As another example, consider a strategy that locates

---

[2]Without the cutoff, R_0 diverges for any positive value of T^*, so the epidemic threshold would be $T^*=0$.

and isolates infecteds compared with a strategy that provides susceptible individuals with protection. Both will be affected by inability to reach everyone. The first strategy has a heterogeneous impact on infectivity, while the second strategy has a heterogeneous impact on susceptibility. If the strategies have the same average impact on T, then the first reduces the probability of an epidemic more while the second reduces its size more. Which strategy is optimal depends on the particular case, and may change with time.

Our results can also be used to identify populations most at risk from epidemics. Populations with low genetic diversity are already known to be at particularly high risk from an outbreak, because the lack of heterogeneity allows the average transmissibility to be higher. However, our results show that, even for a fixed average transmissibility, a population with lower genetic variation will be more severely affected by a disease.

For heterogeneous infectivity but homogeneous susceptibility, Newman [4] anticipated that $\mathcal{A}$ follows from the formulas derived under the assumption of homogeneous T. He did not address the effect on $\mathcal{P}$. We have shown that $\mathcal{A}$ is independent of heterogeneity in infectivity, and so for this special case the prediction is valid. However, it fails if susceptibility is also heterogeneous.

The theory developed here can be generalized in a number of ways. Most simply, we can introduce edge weights to represent some details of the contact between u and v. The same theory will hold, but the calculation of T_o and T_i as in (1) must incorporate the edge weight distribution. We can also introduce correlations between the distributions of $\mathcal{I}$, $\mathcal{S}$, and k in an individual without significant theoretical difficulties, though the conclusions may change. It is more complicated to introduce correlations of $\mathcal{I}$, $\mathcal{S}$, or k between neighbors.

We have considered networks with few short cycles, but true social networks have significant clustering. However, at high transmissibilities, if any neighbors are infected, an epidemic is very likely, so $\mathcal{P}$ is close to the probability that the index case infects any neighbors, and loops may be ignored. At low transmissibilities, loops are not traced out by the infection, and again may be neglected. Loops affect our results only at intermediate transmissibilities. The generating function approach becomes difficult because, even early in an outbreak, infected nodes may have multiple infected neighbors.

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford Science Publications, Oxford, 1991).

[2] L. A. Meyers, M. Newman, and B. Pourbohloul, J. Theor. Biol. **240**, 400 (2006).

[3] L. A. Meyers, Bull., New Ser., Am. Math. Soc. **44**, 63 (2007).

[4] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).

[5] R. R. Kao, L. Danon, D. M. Green, and I. Z. Kiss, Proc. R. Soc. London, Ser. B **274**, 1999 (2006).

[6] N. Madar, T. Kalisky, R. Cohen, D. ben Avraham, and S. Havlin, Eur. Phys. J. B **38**, 269 (2004).

[7] M. A. Serrano and M. Boguñá, Phys. Rev. Lett. **97**, 088701 (2006).

[8] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. **90**, 028701 (2003).

[9] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[10] P. S. Dodds and D. J. Watts, Phys. Rev. Lett. **92**, 218701 (2004).

[11] P. S. Dodds and D. J. Watts, J. Theor. Biol. **232**, 587 (2005).

[12] L. A. Meyers, B. Pourbohloul, M. E. J. Newman, D. M. Skowronski, and R. C. Brunham, J. Theor. Biol. **232**, 71 (2005).

[13] M. E. J. Newman, Phys. Rev. Lett. **95**, 108701 (2005).

[14] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[15] M. B. Hastings, Phys. Rev. Lett. **96**, 148701 (2006).

[16] E. Kenah and J. Robins, e-print arXiv:q-bio/0610057v3.

[17] W. O. Kermack and A. G. McKendrick, Proc. R. Soc. London, Ser. A **115**, 700 (1927).

[18] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Comput. Netw. **33**, 309 (2000).

[19] P. Grassberger, Math. Biosci. **63**, 157 (1983).

[20] J. L. Cardy and P. Grassberger, J. Phys. A **18**, L267 (1985).

[21] H. S. Wilf, *generatingfunctionology*, 3rd ed. (A. K. Peters, Wellesley, MA, 2005).

[22] J. E. Hubbarde, G. Wild, and L. M. Wahl, Genetics (to be published).

[23] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).

[24] M. G. M. Gomes, A. O. Franco, M. C. Gomes, and G. F. Medley, Proc. R. Soc. London, Ser. B **271**, 617 (2004).

[25] G. Chowell, H. Nishiura, and L. M. A. Bettencourt, J. R. Soc., Interface **4**, 155 (2007).

[26] G. Chowell, C. Castillo-Chavez, P. W. Fenimor, C. Kribs-Zaleta, L. Arriola, and J. M. Hyman, Emerg. Infect. Dis. **10**, 1258 (2004).

[27] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).

[28] S. A. Colgate, E. A. Stanley, J. M. Hyman, S. P. Layne, and C. Qualls, Proc. Natl. Acad. Sci. U.S.A. **86**, 4793 (1989).