AMERICAN
Journal of Epidemiology
Formerly AMERICAN JOURNAL OF HYGIENE

VOL. 113 JANUARY, 1981 NO. 1

Reviews and Commentary

INCIDENCE AND PREVALENCE AS USED IN THE ANALYSIS OF THE OCCURRENCE OF NOSOCOMIAL INFECTIONS

FRANK S. RHAME[1] AND WILLIAM D. SUDDERTH[2]

Two recent articles (1, 2) have helped to clarify the proper usage of the terms incidence and prevalence. However, neither paper considered their use in hospital epidemiology. In the analysis of the occurrence of nosocomial infections these terms are used somewhat differently. This paper will discuss the differences and present their rationale. We will then present the mathematical interrelationship of the prevalence rate and incidence rate of nosocomial infection. Finally, we will discuss the practical difficulties and pitfalls in compiling these rates and applying the interrelationship formula.

INCIDENCE RATE

The occurrence of nosocomial infections is most often computed by dividing the number of infections acquired during a given month by the number of patients discharged (or admitted) during that month,

$$I = \text{incidence rate of nosocomial infections for month A} = \frac{\text{number of infections acquired in month A}}{\text{number of patients discharged in month A}} . \quad (1)$$

This parameter is called the infection rate, incidence rate, or, less properly, the incidence. Purists may object to this usage of the word incidence on at least three grounds. The most trivial objection is that the numerator and denominator do not have the same dimension. The numerator is the number of *infections;* the denominator the number of discharged *patients*. On these grounds, equation 1 should be called a ratio. The reason all infections are tallied, rather than infected patients, is that two patients acquiring one infection each is just as undesirable as one patient acquiring two infections. Furthermore, if the number of infected patients were used, an anomaly would arise. If a patient acquired infections in two different months, he/she would be counted twice. If both infections arose

[1] Department of Medicine and Department of Laboratory Medicine and Pathology, University of Minnesota Medical School; University of Minnesota Hospitals and Clinics.
Reprint requests to Dr. Rhame, Infection Control Program, University of Minnesota Hospitals and Clinics, Box 421 Mayo, 420 Delaware Street SE, Minneapolis, MN 55455.
[2] Department of Statistics, University of Minnesota. Research supported by National Science Foundation Grant MCS 77-28424.
The authors thank Drs. Robert W. Haley, David H. Culver, and John V. Bennett for their helpful comments.

during the same month he/she would be counted only once.

A more fundamental problem with the statistic defined by equation 1 is that the numerator and the denominator are not drawn from the same population. A patient who becomes infected in one month may not be discharged until a subsequent month. Such a patient would be counted in the numerator and denominator of different months. A more rigorous way to present incidence data would be to define a cohort (e.g., patients admitted in June) and follow them for instances of nosocomial infection. The numerator of equation 1 would become "instances of infection among June-admitted patients." However, this is not the most useful statistic because outbreaks of infection are more likely to occur at a point in time than within an admission cohort. The rationale for choosing total discharges as the denominator in equation 1 is less secure. The argument against omitting the denominator altogether (producing a true incidence) is that its use adjusts for fluctuations in patient admissions and permits interhospital comparisons. Although substantial variation in the total hospital census is uncommon, equation 1 is also used to produce ward- and service-specific infection rates. In smaller units, census variation is more substantial. The interhospital comparison argument is weak. Meaningful interhospital comparisons are difficult because of differences in hospital population, surveillance methods and definitions of infection. Alternative denominators include patient-days and average daily census (each can be simply computed from the other using the number of days in the survey interval). Use of either does adjust for variations in census and mitigates the following anomaly. For chronic care facilities and services with long durations of hospitalization, equation 1 produces high rates of infection even when few infections are occurring.

The final objection to equation 1 arises because, as used in community epidemiology, incidence rates have a time unit in the denominator (3). For instance, the current incidence rate of measles is about 1.1 cases per 100,000 children per month. In a steady-state hospital, the magnitude of the nosocomial infection incidence rate, as computed by equation 1, is independent of the survey interval. The sense of this parameter is, however, highly analogous to a true incidence rate. The hospital situation is approached differently because of differences in the relative lengths of the survey interval and the sojourn of the population-at-risk in the at-risk status. In community epidemiology the survey interval is usually shorter than the average length of time people are susceptible to the illness. For instance, people are at risk of developing measles for about 10 years and survey intervals of the incidence rate of measles are usually one month or one year. The incidence rate of measles per one month is roughly half that per two months. But how would one compute the incidence rate of measles per century? per millennium? To be meaningful, all the people at risk should be included in the denominator which would then become the number of people having lived during the interval. In a steady-state universe, this parameter would also become time independent. It would be computed in the same way as is the incidence rate defined in equation 1. In the hospital the average length of stay is about eight days and survey intervals range from one month to one year.

Prevalence rate

The prevalence rate is determined at a single point in time. The number of both active and cured nosocomial infections that are or have been present in patients hospitalized on a given day is divided by the number of patients present at the time of the survey,

$$P = \text{prevalence rate of nosocomial infections} = \frac{\text{number of infections (active and cured) having occurred in patients hospitalized at the time of the survey}}{\text{number of patients present at the time of the survey}}.$$

Both active and cured infections are included because of the difficulty in deciding the point at which an infection becomes cured. Some surveys have presented the number of infected patients in the numerator and separate data on multiplicity of infection. In actual practice, the "single point in time" is usually taken to be the interval (generally several hours) during which the survey team visits the ward. Since very few patients acquire a nosocomial infection during their first few hours of hospitalization, this comes very close to a true point prevalence.

Relationship between prevalence rate and incidence rate

Published surveys of the occurrence of nosocomial infection (4–18) have presented both the prevalence and incidence rates (table 1). The incidence rate is more readily conceptualized. When expressed per 100 patients, it is slightly more (because of multiple infections in occasional patients) than the percentage of patients who acquire an infection during their hospitalization. However, the prevalence rate is often determined because it doesn't require the sustained effort needed to produce incidence data. Either appears acceptable in meeting accreditation requirements for surveillance. Since both rates are in use and the easier to obtain is the less desirable, a formula expressing their interrelationship would be useful.

The relationship between P and I is

$$I = P \cdot \frac{LA}{LN - INT} \qquad (2)$$

where LA is the average length of stay of all patients, LN is the average length of stay of patients who acquire one or more nosocomial infections and INT is the average interval between admission and onset of the first nosocomial infection for those patients who acquire one or more nosocomial infections.

The derivation of equation 2 is presented in the appendix for two stochastic models. For both models it is assumed that infections occur independently so that the chance of one patient becoming infected is not dependent on whether or not other patients get infected. To the extent that epidemics and clusters of infection occur, this assumption is unjustified. However, since the bulk of nosocomial infections are endemic, the assumed independence seems close to the truth. Furthermore, the analysis done in the appendix requires only such approximate independence or sufficiently small correlations between the occurrence of infection in different patients.

In the first model it is also assumed, as a first approximation, that patients never acquire more than one nosocomial infection. In the second model, which allows for multiple infections, it is assumed that, for patients who suffered at least one infection, the probability of each subsequent infection does not depend on the number of prior infections. That is, after acquiring a first infection, there is a probability (q in the appendix) of acquiring a second infection. It is assumed that patients acquiring a second infection have the same probability of acquiring a third, and so forth. This assumption is unwarranted insofar as patients who have contracted two infections are likely to be more susceptible to future infection than patients who have acquired one infection. Nevertheless, in two of the three incidence studies presenting multiplicity data (table 2), the frequencies do not deviate greatly from what would be expected under the assumption in question. The deviation that is present is toward in-

TABLE 1

Published hospital-wide surveys of the occurrence of all types of nosocomial infection

Authors	(Reference no.)	Hospital		Survey date	Infections per 100 patients
		Name	Location		
Incidence surveys					
Roy et al.	(4)	Hospital for Sick Children	Toronto, Ontario	Jan. '59–Dec. '59	6.5
McNamara et al.	(5)	U. Kentucky	Lexington, KY	June '65–Aug. '65	6.1
Eickhoff et al.	(6)	6 community	USA	July '65–Dec. '66	3.5
Thoburn et al.	(7)	Johns Hopkins	Baltimore, MD	Oct. '65–March '66	4.0
Gardner and Carles	(8)	Children's	Boston, MA	March '70–Dec. '71	4.6
Mulholland et al.	(9)	U. Pennsylvania	Philadelphia, PA	Oct. '71–Sept. '72	16.7
Wenzel et al.	(10)	U. Virginia	Charlottesville, VA	Sept. '72–Aug. '73	6.7
John	(11)	Moncrief Army	Fort Jackson, SC	July '74–Dec. '75	1.5
Prevalence surveys					
Kislak et al.	(12)	Boston City	Boston, MA	Jan. '64	15.4
Barrett et al.	(13)	Boston City	Boston, MA	Feb. '67	19.7
Adler and Schulman	(14)	Grady Memorial	Atlanta, GA	Oct. '67	13.0
Edwards	(15)	Public Health Service	Staten Island, NY	Feb. '68	8.4
Adler et al.	(16)	Boston City	Boston, MA	Jan. '70	15.0
Moody and Burke	(17)	Latter Day Saints	Salt Lake City, UT	Jan. '71	9.2
Britt et al.	(18)	18 inter-mountain	Rocky Mountain States	Oct. '72–Feb. '73	7.6
Mulholland et al.	(9)	U. Pennsylvania	Philadelphia, PA	June '73	14.4

TABLE 2

Multiplicity of infection; conditional frequencies of subsequent infections

	Incidence surveys			Prevalence surveys			
	(4)	(*)	(8)	(12)	(16)	(17)	(18)
Total patients studied	17,836	87,708	12,209	928	709	566	525
No. of patients with							
1 infection	972	1,253	920	108	72	44	37
2 infections	76	141	55	16	9	4	1
3 infections	8	26	31	1	4		
4 infections	2	1	15				
5 infections	1		1				
Observed conditional frequencies							
≥1 infection	0.059	0.016	0.032	0.135	0.120	0.085	0.072
≥2 infections given 1 infection	0.082	0.118	0.264	0.136	0.153	{0.083}	{0.026}
≥3 infections given 2 infections	0.126	0.161	0.461	{0.059}	{0.308}		
≥4 infections given 3 infections	{0.273}†	{0.037}	0.340				
≥5 infections given 4 infections	{0.333}		{0.062}				

* Unpublished data observed in a continuation of the study described by Eickhoff et al. (6). The infections occurred between July 1968 and June 1969.

† Frequencies in braces were computed from <10 infections.

creasing likelihood of infection as more infections occur. Some of this increase in the occurrence of multiple infections is probably artifactual. Incidence surveillance is usually done by reviewing a subset of the charts of currently hospitalized patients. Various strategies may be used to choose the charts of patients with increased likelihood of having had a nosocomial infection. For instance, surveillance personnel may examine the charts of patients who are febrile and/or have had cultures obtained. This sample is probably biased toward multiple infection patients. Furthermore, subsequent infections are more likely to be identified by surveillance personnel than first infections because the patient's chart must be carefully examined in the course of determining and assessing the data bearing on the first infection. In support of this, the conditional frequencies are more nearly similar for the survey of Roy et al. (4) which involved a uniform chart review. Otherwise, these models are very general. Specifically, no presumptions about day-specific infection rates are required.

In the prevalence surveys (table 2), the assumption seems to be quite consistent with the data partly because very few patients were observed to acquire two infections and even fewer to acquire three. In any case, the defects in equation 2 caused by the assumption are minor. With an I of 0.05, which is typical, the contributions of infections after the second are quite small.

The only additional assumption is that the expected intervals between infections are identical. Since fourth infections are rare enough to be negligible, the substance of the assumption is that the average interval between the first and second infection is the same as that between the second and third infection. While we know of no data bearing on this point, we also know of no better presumption.

Examination of equation 2 allows a more precise statement of the basis of the generally observed greater magnitude of the prevalence rate compared to the incidence rate. Specifically, $(LN - INT)$ must be greater than LA. That is, the length of time patients are hospitalized after acquiring nosocomial infection must be longer than the average length of admission.

In the epidemiology of community acquired disease, the relationship of incidence and prevalence rates is described by

$$\text{Prevalence Rate} = \text{Incidence Rate} \times \text{Duration of Illness of Active Infection.}$$

Equation 2, rearranged, is similar

$$P = \frac{I}{LA} \cdot (LN - INT).$$

P is the prevalence of both actively infected and cured patients; $(LN - INT)$ is the duration of that combined condition. The "true" incidence rate (3) as used in community epidemiology is analogous to the term I/LA. For instance, in a hospital with LA = eight days and a nosocomial infection incidence rate of five infections per 100 discharged patients, the "true" incidence rate of nosocomial infections would be 0.625 infections per 100 patient days. When LA is known, data from surveys using patient-day denominators may be converted to the incidence rate defined in equation 2.

Application within the hospital

Both incidence and prevalence surveys fail to identify inadequately documented or unrecognized infections. Beyond this, the two surveillance methods tend to obtain different biased samples of the infections which can be documented by a complete, post discharge chart review. Prevalence surveys tend to miss infections with a greater lag between onset and documentation. A patient with fever occurring on the day before the survey who has a blood culture drawn which does not turn positive until the day after the survey will not be recorded as infected. They are also biased toward patients who have longer

than average durations of stay. Incidence surveillance is usually carried out so that infections can be identified soon after they occur rather than in record rooms on the charts of discharged patients. Generally, surveillance personnel review only a selected portion of charts of currently hospitalized patients. Patients are selected using clues (e.g., those patients from whom a culture is obtained) which yield a population with a high probability of being infected. This method generally detects those infections which tend to produce the clues which lead to chart review. To the extent that prevalence and incidence surveys identify different biased samples of all infected patients, equation 2 is invalid.

LA is readily available in most hospitals. It is usually computed by dividing the average daily census by the average daily admissions. Unfortunately, this computation produces a systematic exaggeration of *LA* because most hospitals generate an "average daily census" rather than an "average instantaneous census." In the former are included all patients hospitalized during the midnight-to-midnight 24-hour period, even those discharged just after the day begins or admitted just before midnight. In a hospital with a true *LA* of eight days, where all admissions and discharges occur at noon and the average daily census includes all patients hospitalized from midnight, the computed *LA* would be nine days (12.5 per cent high). This problem may be partially compensated for by including all the patients in the prevalence who are on the ward at any time of the day of the survey.

The average length of stay of patients acquiring at least one nosocomial infection (*LN*) and their interval from admission to onset of infection (*INT*) may be closely approximated during a prevalence survey by tallying *INT* and attaching to the charts of infected patients a form to be completed with the discharge date and mailed to survey personnel. This method may not be exact because the sample of patients present may be biased toward a longer staying subgroup of patients acquiring a nosocomial infection. *LN* would be artifactually higher; bias in *INT* in such a sample is unknown. A more rigorous *INT* and *LN* can be obtained by surveying a cohort of patients (say, those admitted during several randomly selected days).

In the future, published surveys presenting prevalence data should be accompanied by the simply obtainable data which will enable the conversion to incidence data. *LN* and *INT* should be determined by the methods described in the previous paragraph. *LA* may be approximated from the average daily census and the average daily admissions for the month during which the survey was performed. Obtaining the average length of stay for the patients present on the day of the survey would probably not be a good approximation of *LA;* this patient sample is biased toward longer staying patients.

Note: Since the submission of this manuscript, we have become aware of a paper (21) describing the relationship between incidence and prevalence.

References

1. Elandt-Johnson RC. Definition of rates: some remarks on their use and misuse. Am J Epidemiol 1975;102:267-71.
2. Friedman GD. Medical usage and abusage: "prevalence" and "incidence." Ann Intern Med 1976;84:502-4.
3. MacMahon B, Pugh TF. Epidemiology. Boston: Little Brown and Co, 1970.
4. Roy TE, McDonald S, Patrick ML, et al. A survey of hospital infection in a pediatric hospital: Part I. Description of hospital, organization of survey, population studied and some general findings. Can Med Assoc J 1962;87:531-8.
5. McNamara MJ, Hill MC, Balows A, et al. A study of the bacteriologic patterns of hospital infections. Ann Intern Med 1967;66:480-8.
6. Eickhoff TC, Brachman PS, Bennett JV, et al. Surveillance of nosocomial infections in community hospitals. I. Surveillance methods, effectiveness and initial results. J Infect Dis 1969; 120:305-16.
7. Thoburn R, Fekety FR Jr, Cluff LE, et al. Infections acquired by hospitalized patients: an analysis of the overall problem. Arch Intern Med 1968;121:1-10.

8. Gardner P, Carles DG. Infections acquired in a pediatric hospital. J Pediatr 1972;81:1205-10.
9. Mulholland SG, Creed J, Dieraug LA, et al. Analysis and significance of nosocomial infection rates. Ann Surg 1974;180:827-30.
10. Wenzel RP, Osterman CA, Hunting KJ, et al. Hospital-acquired infections. I. Surveillance at a university hospital. Am J Epidemiol 1976; 103:251-60.
11. John JF Jr. Nosocomial infection rates at a general Army hospital. Am J Surg 1977;134:381-4.
12. Kislak JW, Eickhoff TC, Finland M. Hospital-acquired infections and antibiotic usage in the Boston City Hospital—January, 1964. N Engl J Med 1964;271:834-5.
13. Barrett FF, Casey JI, Finland M. Infections and antibiotic use among patients at Boston City Hospital, February, 1967. N Engl J Med 1968; 278:5-8.
14. Adler JL, Schulman JA. Nosocomial infection and antibiotic usage at Grady Memorial Hospital: a prevalence survey. South Med J 1970; 63:102-5.
15. Edwards LD. Infections and use of antimicrobials in an 800-bed hospital. Public Health Rep 1969;84:451-7.
16. Adler JL, Burke JP, Finland M. Infection and antibiotic use at Boston City Hospital, January, 1970. Arch Intern Med 1971;127:460-5.
17. Moody ML, Burke JP. Infections and antibiotic use in a large private hospital, January, 1971. Arch Intern Med 1972;130:261-6.
18. Britt MR, Burke JP, Nordquist AG, et al. Infection control in small hospitals. Prevalence surveys in 18 institutions. JAMA 1976;236:1700-3.
19. Feller W. Introduction to probability theory and its applications. Volume 2. New York: John Wiley & Sons, 1971.
20. Chung KL. A course in probability theory. New York: Harcourt, Brace and World, 1968.
21. Freeman J, Hutchison GB. Prevalence, incidence and duration. Am J Epidemiol 1980; 112:707-23.

APPENDIX: TWO STOCHASTIC MODELS FOR NOSOCOMIAL INFECTIONS IN A LARGE HOSPITAL

The epidemiologic assumptions underlying these models and the definitions of P and I have been discussed in the main text. The simpler model is based on the presumption that patients do not acquire more than one nosocomial infection.

Let $X_1, X_2, \ldots X_n, \ldots$ be random variables corresponding to the time spent in the hospital by successive occupants of a single hospital bed. The X_n's are assumed to be independent and have the same distribution.

Let i = the probability that the nth patient in the bed contracts an infection. (3)

i is assumed to be the same for all patients. Let Z_n^0 be a variable corresponding to the time spent in the hospital by the nth patient prior to the occurrence of any infection and let Z_n^1 be the time spent by the patient after an infection occurs. $Z_n^1 = 0$ if patient n is never infected.

Then $X_n = Z_n^0 + Z_n^1$. (4)

Now take F_n^0 to be the event that the nth patient does not become infected and F_n^1 to be the event that one infection is acquired. By equation 4, the following relationship between expected lengths of stay may be expressed

$$E(X_n|F_n^1) = E(Z_n^0|F_n^1) + E(Z_n^1|F_n^1). \quad (5)$$

Next suppose that the process corresponding to a single bed is examined after it has been in operation for a large time t, and define

p = probability that the patient in the bed at time t has been infected.

Then, under very mild assumptions on the distribution of X which doubtless hold in practice, the renewal theorem (see reference 19, Theorems XI.1.1 and XI.1.2 and Example XI. 8a) implies that

$$p \approx \frac{\text{expected time in hospital for any patient after being infected}}{\text{expected time in hospital for any patient}}$$

$$= \frac{iE(Z_n^1 | F_n^1)}{EX_n}. \quad (6)$$

Solving equation 5 for $E(Z_n^1 | F_n^1)$, substituting in equation 6, and rearranging yields

$$i = p \cdot \frac{EX_n}{E(X_n | F_n^1) - E(Z_n^0 | F_n^1)}. \quad (7)$$

Now assume that the hospital contains a large number of beds. By definition, P calculated at time t is just the proportion of patients in the hospital who have been infected. Assume also that the random variables associated with different beds are independent or at least have small correlation. Then, by the law of large numbers,

$$P \approx p. \quad (8)$$

It is also true that if I is the incidence rate for a month beginning at time t, then

$$I \approx i. \quad (9)$$

To verify equation 9, calculate as follows:

$$I = \frac{\text{number of infections in the given month}}{\text{number of patients admitted in the given month}}$$

$$\approx \frac{\text{expected number of infections}}{\text{expected number of patients admitted}}$$

$$\approx \frac{\frac{\text{length of the month}}{EY_m} \cdot Q}{\frac{\text{length of the month}}{EX_n} \cdot Q}$$

$$= \frac{EX_n}{EY_m}, \quad (10)$$

where Y_1 is the time of dismissal of the first infected patient and where Y_m is the time from the dismissal of the $(m-1)$st infected patient until the time of dismissal of the mth infected patient. The successive lines in equation 10 are by definition, by the law of large numbers, by the renewal theorem, and obvious (Q is the number of beds in the hospital), respectively. If S is the number of patients up to and including the first to be infected, then S is a geometric random variable with expectation i^{-1}. Y_1 can be written in the form

$$Y_1 = \sum_{n=1}^{S} X_n.$$

By an equation due to Wald (see reference 20, Theorem 5.5.3),

$$EY_1 = (ES)(EX_1) = i^{-1} EX_1. \quad (11)$$

Because the X_n's and the Y_m's all have the same probability distribution, equation 11 may be generalized to

$$i = \frac{EX_n}{EY_m}. \quad (12)$$

By equations 10 and 12, equation 9 is established. By the law of large numbers again, the expected values in equation 7 are approximately equal to the respective average values. Thus by equations 7, 8, and 9

$$I \approx P \cdot \frac{LA}{LN - INT}.$$

In the second, more complex model, patients may acquire any number of infections. Corresponding to equation 4, there is now the equality

$$X_n = Z_n^0 + Z_n^1 + Z_n^2 + \cdots$$

where Z_n^0 is defined as before, and, for $k \geq 1$, Z_n^k is the time spent by the nth patient from the onset of the kth infection until the onset of the $k+1$st or dismissal from the hospital. Z_n^k is set equal to zero if there is no kth infection. Take F_n^k to be the event that the nth patient develops exactly k infections. Generalizing equation 5, the relationship between expected lengths of stay, given that F_n^k occurs, may be expressed

$$E(X_n|F_n^k) = E(Z_n^0|F_n^k) + E(Z_n^1|F_n^k) + \cdots + E(Z_n^k|F_n^k).$$

Now let i_k = probability that any patient contracts exactly k infections, and p_k = probability that the patient in bed at time t has contracted k infections. Again, by the renewal theorem, for t sufficiently large,

$$p_k \approx \frac{\text{expected time in hospital for any patient during which the patient has exactly } k \text{ infections}}{\text{expected time in hospital for any patient}}$$

$$= \frac{\sum_{l=k}^{\infty} i_l E(Z_n^k|F_n^l)}{EX_n}. \tag{13}$$

Equation 13 may be simplified by the assumptions discussed in the text of the article. The probability q of developing one subsequent infection, given that at least one infection has occurred, is assumed to be independent of the number of prior infections. This assumption gives a geometric distribution for the i_k's except possibly at the first term i_0.

$$i_k = q^{k-1} i_1 \text{ for } k = 1,2,\ldots. \tag{14}$$

It is also assumed that the expected time to a subsequent infection after a first or subsequent infection is a constant C regardless of the number of prior infections

$$E(Z_n^j|F_n^k) = C \text{ for all } n \text{ and } 1 \leq j \leq k. \tag{15}$$

Let A_n be the event that the nth patient has at least one infection and let

$$T_n = Z_n^1 + Z_n^2 + \cdots \tag{16}$$

be the total length of stay after the first infection occurs. Then

$$X_n = Z_n^0 + T_n \quad \text{and, consequently,}$$
$$E(X_n|A_n) = E(Z_n^0|A_n) + E(T_n|A_n). \tag{17}$$

Using equations 14, 15, and 16, one can calculate the expected value of T_n given A_n.

$$E(T_n|A_n) = \sum_{k=1}^{\infty} E(T_n|F_n^k) P(F_n^k|A_n)$$

$$= \sum_{k=1}^{\infty} (kC) \quad (q^{k-1}(1-q))$$

$$= \frac{C}{1-q} \quad . \tag{18}$$

Using equation 13 for $k \geq 1$ and equations 14, 15, and 18

$$p_k \approx \frac{\sum_{l=k}^{\infty} q^{l-1} i_1 C}{EX_n}$$

$$= \frac{C i_1 q^{k-1}}{(EX_n)(1-q)}$$

$$= \frac{i_k E(T_n | A_n)}{EX_n} \quad . \tag{19}$$

Thus, for $k \geq 1$, the p_k's are proportional to the i_k's and, hence, also follow a geometric distribution. In fact, the assumption of equation 14 that the i_k's are geometric can be shown in the presence of equation 15 to be equivalent to the same assumption about the p_k's. This approximate equality is the key to the relationship between the quantities I and P in the model under consideration. The prevalence P is, by the law of large numbers, approximately the expected number of infections having been contracted by a patient at time t. So, for large t,

$$P \approx p_1 + 2p_2 + 3p_3 + \ldots \tag{20}$$

The incidence I can be written in the form

$$I = I_1 + 2I_2 + 3I_3 + \ldots$$

where

$$I_k = \frac{\text{number of patients to have exactly } k \text{ infections in a given month}}{\text{number of patients admitted in the given month}} \quad .$$

The same argument given for equation 9 also shows that, for every k, $I_k \approx i_k$ and that

$$I \approx i_1 + 2i_2 + 3i_3 + \ldots \tag{21}$$

From equations 19, 20, and 21

$$I \approx P \cdot \frac{EX_n}{E(T_n | A_n)} \quad . \tag{22}$$

Solving equation 17 for $E(T_n | A_n)$ and substituting in equation 22 yields

$$I \approx P \cdot \frac{EX_n}{E(X_n | A_n) - E(Z_n^0 | A_n)} \quad .$$

The expected values may be approximated by their averages as before

$$I \approx P \cdot \frac{LA}{LN - INT} \quad .$$