Laboratory and Epidemiology Communications

Parameters That Characterize Different Food-Poisoning Outbreaks

Hiroshi Yoshikura*

National Institute of Infectious Diseases, Tokyo 162-8640, Japan

Communicated by Haruo Watanabe

(Accepted January 31, 2012)

Various microbes and toxins cause food poisoning under different settings. The Statistics of Food Poisoning, Ministry of Health, Labour and Welfare of Japan, provides food poisoning data in Excel format (http://www.mhlw.go.jp/topics/syokuchu/04.html#4–3). This study shows that food poisoning caused by different microbes has different frequency distribution of the number of patients per incident, and that the distribution pattern of enterohemorrhagic *Escherichia coli* (EHEC) is completely different from those of other food poisoning microbes, including pathogenic *E. coli* other than EHEC (non-EHEC *E. coli*). Here, it also shows that the frequency distribution of the number of patients per incident in cases of food poisoning by a natural toxin follows the power law.

Microbial food poisoning: Fig. 1 shows procedures for obtaining the cumulative frequency distribution by using food poisoning by *Clostridium perfringens* as an example. *C. perfringens* was selected from the column "responsible agents" in the data file titled "the list of the individual food poisoning incidents." The step was followed by selection of the "patient number per incident" column; the elements in the column were rearranged in an ascending order and copied to a new Excel file. Only those incidents with 2 patients or more were chosen, because very often incidents involving only 1 person are not reported as outbreaks. Ranking number was allocated to "patient number per incident" as shown on the left side of Fig. 1. The patient number per incident and ranking number were plotted on the X and Y axes, respectively. The X axis was then converted to a log scale. The completed plot is shown on the right side of Fig. 1. Although the cumulative frequency distribution curve traces the uppermost points, in this study, the plot itself will be called as such for the purpose of convenience.

The median (M) of patient number per incident is X at Y = T/2, where T is total number of incidents. The upper quartile boundary (UQB) and the lower quartile boundary (LQB) are X at Y = 3T/4 and at Y = T/4, respectively. Interquartile range (IQR) is defined as log UQB–log LQB, which is equal to log UQB/LQB (see tabulation on the left side of Fig. 1). Terms, M, LQB, HQB, and IQR, were used as described in a textbook by Aitken et al. (1).

Panels A, B, and C in Fig. 2 show cumulative fre-

quency distribution curves for *Staphylococcus aureus* (A), *Salmonella* (B), and non-EHEC *E. coli* (C, plot in diamonds), which are sigmoidal and have point symmetry with the center near the coordinates (M, T/2). All the food poisoning-causing microbes examined so far, except EHEC (C, plot in squares), had similar frequency distribution. The bar graph (panels A and B) showed near log-normal distribution, and M value was within the range of the peak bar, i.e., the frequency distribution was near log-normal.

Frequency distribution of EHEC was completely different from that of the other microbes (panel C); the EHEC distribution showed a left-upward convex shape, and the bar graph resembled a staircase descending towards the right (bar graph in solid lines). On the other hand, the cumulative frequency distribution for non-EHEC *E. coli* was sigmoidal. The M value of non-EHEC *E. coli* was within the peak bar (bar graph in dotted lines), while that of EHEC was not. These findings suggest that the epidemic pattern for outbreaks caused by EHEC is different from that for outbreaks caused by non-EHEC *E. coli* and other microbes.

Table 1 summarizes M and IQR values for different microbes and different facilities. For *Norovirus* (NoV), *Campylobacter jejuni/coli*, and *Salmonella*, M and IQR values did not change much over a period of 3 years (2008–2010). The respective values were 18–23 and 0.571–0.621 for NoV, 6–7 and 0.301–0.477 for *C. jejuni/coli*, 13–15 and 0.663–0.745 for *Salmonella*, 11–14 and 0.427–0.670 for *S. aureus*, 13–15 and 0.634–0.813 for *Vibrio parahaemolyticus*, 6 and 0.635 for *Bacillus cereus*, 41 and 0.56 for *C. perfringens*, and 22 and 0.533 for non-EHEC E. coli (Table 1A). Since the M value reflects the outbreak size, each microbe had its own outbreak size.

Food preparing facilities: Table 1B shows a comparison of different food preparing facilities for all the agents (the first three rows) and for NoV alone (the next five rows). The M values for hotels, restaurants, and caterers were 23, 10, and 42, respectively, and those for NoV-hotel, NoV-restaurant, and NoV-caterer combinations were 29, 16, and 46, respectively. However, the M values for caterer-non-NoV combination and that of NoV-non-caterer combination were 24 and 18, respectively, and were almost half of that for the NoV-caterer combination (M = 46). Thus, the combination of caterers and NoV, neither caterers nor NoV alone, resulted in the high M value (M = 46). The rest of the table shows the comparison of food poisoning caused by different microbes in restaurants. Even in the same setting, the M values were different for different microbes; the M values were 34 for *C. perfringens*; 13 for

---

*Corresponding author: Mailing address: Department of Food Safety, Ministry of Health, Labour and Welfare, 1-2-2 Kasumigaseki, Chiyoda-ku, Toyo 100-8916, Japan. E-mail: yoshikura-hiroshi@mhlw.go.jp

X axis — Patients/incident | Y axis — Ranking number

X axis Patients/incident	Y axis Ranking number	
4	1	
5	2	
6	3	
7	4	
7	5	
17	19	LQB
19	20	T/4
21	21	
21	22	
21	23	
21	24	
22	25	
26	26	
28	27	
30	28	
30	29	
30	30	
31	31	
32	32	
33	33	
34	34	
36	35	
37	36	
38	37	
40	38	
41	39	Median, T/2
42	40	
43	41	
48	42	
49	43	
49	44	
50	45	
53	46	
53	47	
53	48	
54	49	
55	50	
56	51	
57	52	
57	53	
60	54	
61	55	
63	56	
64	57	
66	58	
69	59	HQB, 3T/4
69	60	
72	61	
397	77	
645	78	Total, T

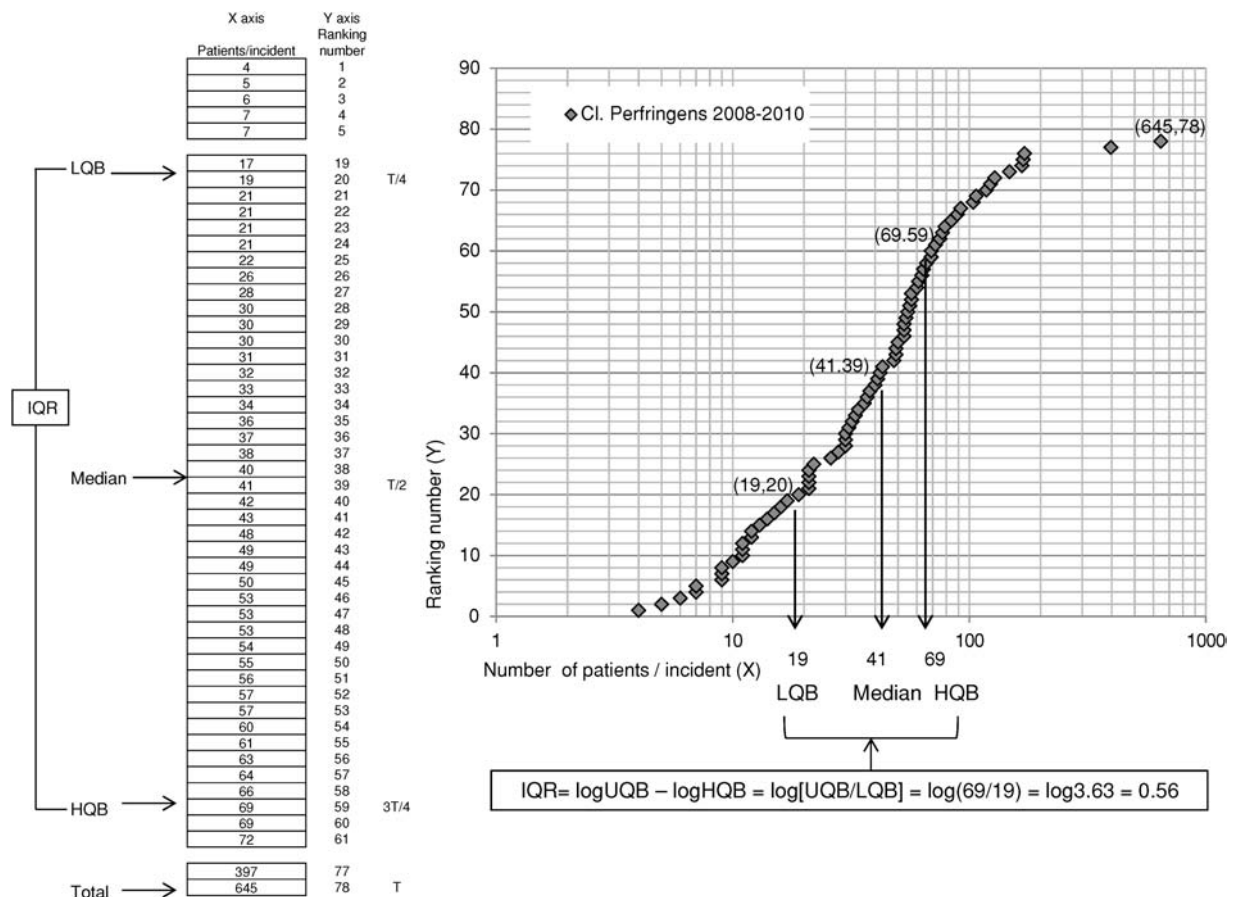$$IQR = \log UQB - \log HQB = \log[UQB/LQB] = \log(69/19) = \log 3.63 = 0.56$$

Fig. 1. Procedures for producing a cumulative frequency distribution curve by using Microsoft Office Excel 2007. The completed worksheet is shown on the left side, and the obtained cumulative frequency distribution curve is shown on the right, where the patient number/incident is plotted on the X axis (log scale) and the ranking number (= cumulative number of incidents) is plotted on the Y axis (ordinary scale). The data file of the Statistics of Food Poisoning contains information on place, date, description of suspected foods/dishes, identified causative agents including natural toxins, food preparing facilities, number of consumers, number of patients, and number of deaths. The Excel format of the available data sets is very convenient for fast and accurate construction of the relevant figures and tables.

V. parahaemolyticus, *Salmonella*, and *S. aureus*; and 6 for *B. cereus* and *C. jejuni/coli*.

Implicated foods: Although most incidents could not be traced back to particular foods or dishes (for example, about 720 of 990 NoV-related incidents that occurred in 2008–2010 could not be traced back to particular foods), the ''implicated foods'' list in Statistics of Food Poisoning contained limited but invaluable information.

– Although both *B. cereus* and *C. perfringens* share a common character in forming spores, the M value for *B. cereus* was 6 and that for *C. perfringens* was 41. Fried rice and noodles were implicated in 18 of 24 cases of food poisoning by *B. cereus*; however, these foods were not implicated in food poisoning by *C. perfringens*. Meat dishes including curry and rice, were implicated in 18 of 35 cases of food poisoning by *C. perfringens* (M = 41), but were implicated in only one case of food poisoning by *B. cereus*. The different M values for the two bacteria could be attributed to the entirely different kinds of implicated foods, which in turn could be further attributed to the different habitats of *B. cereus* (aerobic) and *C. perfringens* (anaerobic).

– In 192 cases of *C. jejuni/coli* (M = 6–7) food poisoning, raw or undercooked meat, mainly of chicken and cattle origin, was implicated in 64 incidents. The low M value could be due to the peculiar eating style, i.e., eating of raw or undercooked meat, which occurs in limited places with a limited number of clients.

– In cases of food poisoning by *V. parahaemolyticus* (M = 13–15), although 50 of 56 incidents could not be traced back to foods, the remaining 6 were all traced back to fishes, which can be explained by the microbe's marine habitat.

– In cases of NoV food poisoning, implicated foods were identified in about 270 incidents. Among the implicated foods, raw or undercooked oysters and other bivalves were implicated in as many as 102 incidents. However, the M value for such cases was 11, which was half of that for all the NoV cases (M = 18–23) and far lower than that for NoV-caterer combination (M = 46). Among the total of 151 cases of caterer-mediated NoV food poisoning, however, only six could be traced back to foods, and only four of six sushi-related incidents were linked to bivalves. Therefore, bivalves alone were probably not the
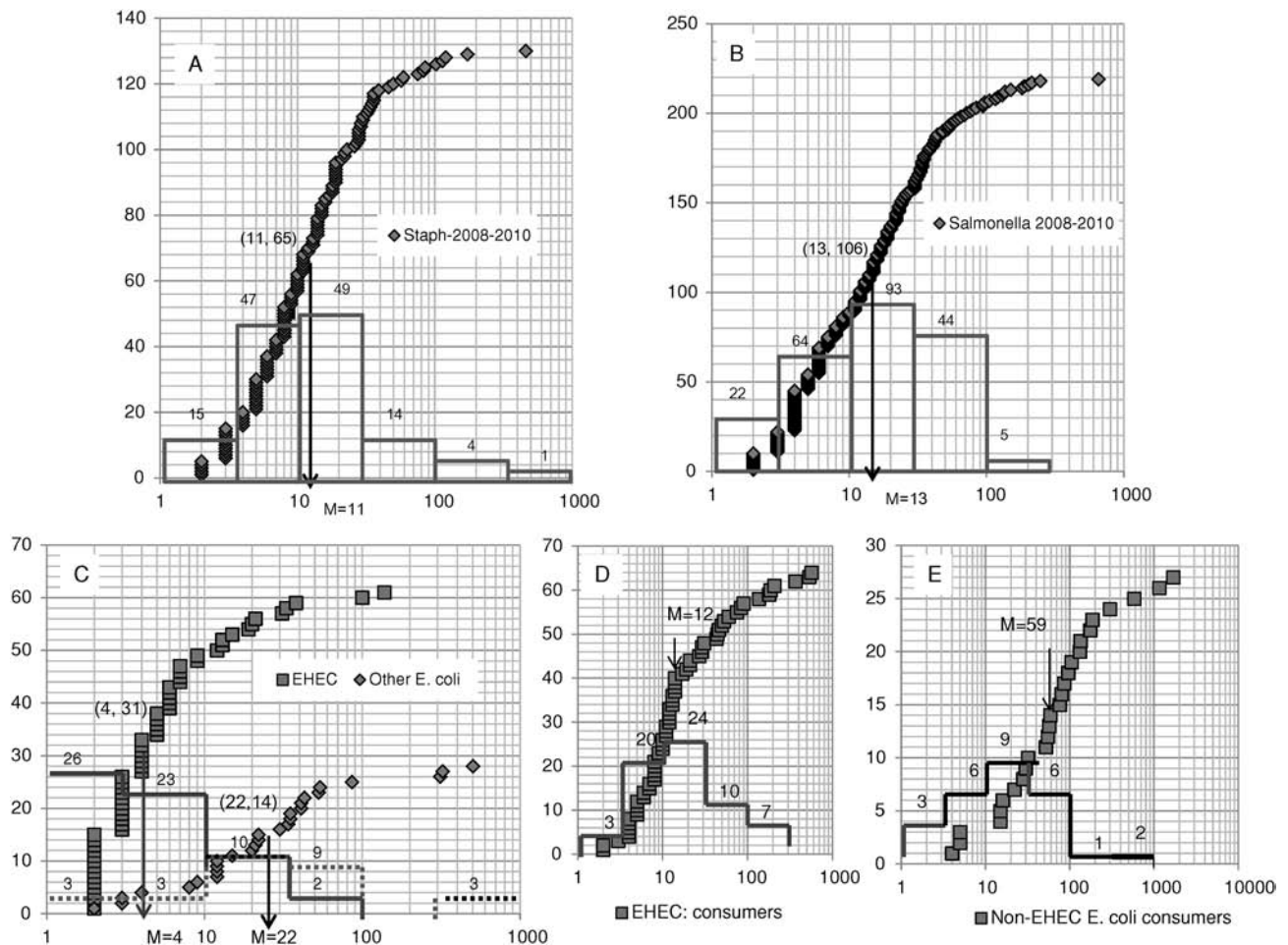
Fig. 2. Examples of cumulative frequency distribution curves. The patient number/incident is plotted on the X axis (log scale), and the ranking number is plotted on the Y axis (ordinary scale). The coordinates within the figures indicate the origins of the arrows that show the position of M on the X axis. (A) *S. aureus*, (B) *Salmonella*, (C) EHEC and non-EHEC *E. coli*, (D) cumulative frequency distribution of consumers/incident for EHEC, (E) cumulative frequency distribution of consumers/incident for non-EHEC *E. coli*. The width of the bar is 1/2 log 10 or log 3.16; number of incidents is shown above each bar. For panel C, left to right, number of incidents for each bar is 26, 23, 10, and 2 for EHEC (solid line), and 3, 3, 10, 9, 0, and 3 for non-EHEC *E. coli* (dotted line).

main cause of the high M value for NoV (M = 18–23) or NoV-caterer combination (M = 46).

Relationship between the number of patients and the number of consumers: Since the number of patients is always less than the number of consumers, the frequency distribution of patients is not independent of that of consumers and should be taken into account when evaluating the correlation coefficients (CC) between the number of patients and consumers. CC was calculated using Excel CORREL. The right-most column of Table 1 lists CC between the number of patients/incident and the number of consumers/incident (CC). Correlation was high for non-EHEC *E. coli* (CC = 0.956) and moderately high for *S. aureus* (CC = 0.895), *C. perfringens* (CC = 0.863), *B. cereus* (CC = 0.821), and NoV (CC = 0.86). For other microbes, such as *C. jejuni* (CC = 0.68), *V. parahaemolyticus* (CC = 0.546), *Salmonella* (CC = 0.675), and EHEC (CC = 0.639), and for facilities such as hotels (CC = 0.673), restaurants (CC = 0.678), and caterers (CC = 0.644), the correlation was rather marginal.

Panels D and E in Fig. 2, respectively, show cumula-tive frequency distribution of number of consumers per food poisoning incident by EHEC and non-EHEC *E. coli*, which were very similar and sigmoidal, i.e., log-normal distribution. However, the frequency distribution was, as already indicated, log-normal for non-EHEC but not for EHEC (see panel C). The CC values was 0.956 for non-EHEC *E. coli* and 0.639 for EHEC. It was suggested that (i) log-normal distribution of con-sumers/incident does not necessarily result in log-nor-mal distribution of patients/incident, (ii) for cases with marginal CC values (around 0.64), similar sigmoid fre-quency distribution curves for number of consumers/ incident and number of patients/incident will not neces-sarily mean that they are causally related (this is rather obvious if we recall that school record and height both showing normal distribution are unrelated), and (iii) a higher value of CC may indicate strong correlation be-tween frequency distribution of consumers and patients.

Short summary: The above observations suggest that the apparently simple parameters M and IQR are deter-mined by complex interactions of different factors, e.g., microbes, foods/dishes, food-preparing facilities, num-

Table 1. M and IQR for different microbes, facilities, and their combination

Table A	Pathogen	Year	T/2	M	LQB	UQB	IQR	CLLP	CC
	NoV	2008	151.5	23	11	46	0.621	(749, 303)	
	NoV	2009	144	21	11	41	0.571	(636, 288)	
	NoV	2010	199.5	18	10	37	0.568	(197, 399	
	NoV	2008–2010	495	20	11	41	0.572	(197, 990)	0.86
	C. jejuni/coli	2008	151.5	6	4	12	0.477	(110, 303)	
	C. jejuni/coli	2009	105.5	7	5	10	0.301	(155, 211)	
	C. jejuni/coli	2010	121	6	4	9	0.352	(49, 242)	
	C. jejuni/coli	2008–2010	378	6	4	11	0.439	(155, 756)	0.68
	V. parahaemolyticus	2010	16.5	15	4	26	0.813	(54, 32)	
	V. parahaemolyticus	2008–2010	28	13	6	26	0.634	(67. 56)	0.546
	Salmonella	2008	43.5	15	5	26	0.716	(48, 74)	
	Salmonella	2009	31.5	14	7	30	0.633	(202, 63)	
	Salmonella	2010	34.5	13	6	34	0.754	(654, 69)	
	Salmonella	2008–2010	105.5	13	5	26	0.716	(654, 206)	0.675
	S. aureus	2008	28	11	5	19	0.56	(460, 56)	
	S. aureus	2009	20.5	11	6	19	0.427	(102, 41)	
	S. aureus	2010	16.5	14	6	28	0.67	(173, 33)	
	S. aureus	2008–2010	65	11	6	22	0.565	(460, 130)	0.895
	B. cereus	2008–2010	24.5	6	3	13	0.635	(40, 49)	0.821
	C. perfringens	2008–2010	39	41	19	69	0.56	(645, 78)	0.863
	E. coli other than EHEC	2008–2010	14	22	12	41	0.533	(503, 28)	0.956
	EHEC	2008–2010	[30.5]	[4]	[2]	[7]	[0.544]	(138, 61)	0.639

Table B	Pathogen	Facility	T/2	M	LQ	UQ	IQR	CLLP	CC
	all	Hotel	123	23	14	47	0.526	(636, 246)	0.673
	all	Restaurant	912	10	6	20	0.522	(321, 1273)	0.678
	all	Caterer	70.5	42	21	96	0.66	(1197, 141)	0.644
	NoV	Hotel	50	29	14	70	0.699	(636, 149)	0.949
	NoV	Other than caterer	452	18	10	37	0.568	(636, 904)	0.776
	NoV	Caterer	43	46	26	98	0.576	(1197, 86)	0.924
	Other than NoV	Caterer	27.5	24	12	92	0.89	(655, 55)	0.655
	NoV	Restaurant	334	16	9	32	0.551	(255, 668)	0.715
	C. jejuni/coli	Restaurant	294.5	6	4	10	0.398	(51, 589)	0.55
	V. parahaemolyticus	Restaurant	17	13	8	26	0.512	(130, 34)	0.452
	S. aureus	Restaurant	29.5	13	6	19	0.501	(120, 59)	0.682
	B. cereus	Restaurant	16	6	3	13	0.635	(40, 32)	0.994
	C. perfringens	Restaurant	19	34	12	69	0.76	(171, 38)	0.802

Table A: M and IQR for different microbes. Table B: M and IQR for different combinations of microbes and facilities (figures are for the total of 3 years, 2008–2010). CLPP: the coordinates of the last plot point in the cumulative frequency distribution curve, which indicates the highest number of patients/incident on the left and the total number of incidents on the right. CC: correlation coefficient (calculated by Excel CORREL) between the number of patients and number of consumers for incidents in which both information data are available. For EHEC, figures other than for CLPP and CC are bracketed as they are meaningful only in normal distribution. T, total number of incidents; M, median; LQ, lower quartile; UQ, upper quartile; IQR, interquartile range.

ber of consumers, and probably distribution of microbes in foods, which is known to follow log-normal distribution (2). It is interesting to note that, in spite of such complexity, the parameters remain quite stable over time.

Parameters M and IQR combined with frequency distribution of the number of consumers/incident and its correlation with the number of patients/incident (CC) could be used as tools for a more detailed analysis of food poisoning events. It would be interesting to know, e.g., how implementation of a specific hygienic measure can influence these parameters.

Food poisoning by natural toxins: The data file used for the analysis of microbial food poisoning contains data for natural toxins. Most plant and animal toxins found in the file are from the poisonous mushrooms and plants and from poisonous fish and shell fish, respectively. Fig. 3 shows the frequency distribution of the number of patients per incident. The plot is in log-log, and the number of patients/incident is plotted on the X axis and frequency is plotted on the Y axis. Here, the single-patient cases were included, because in princi-
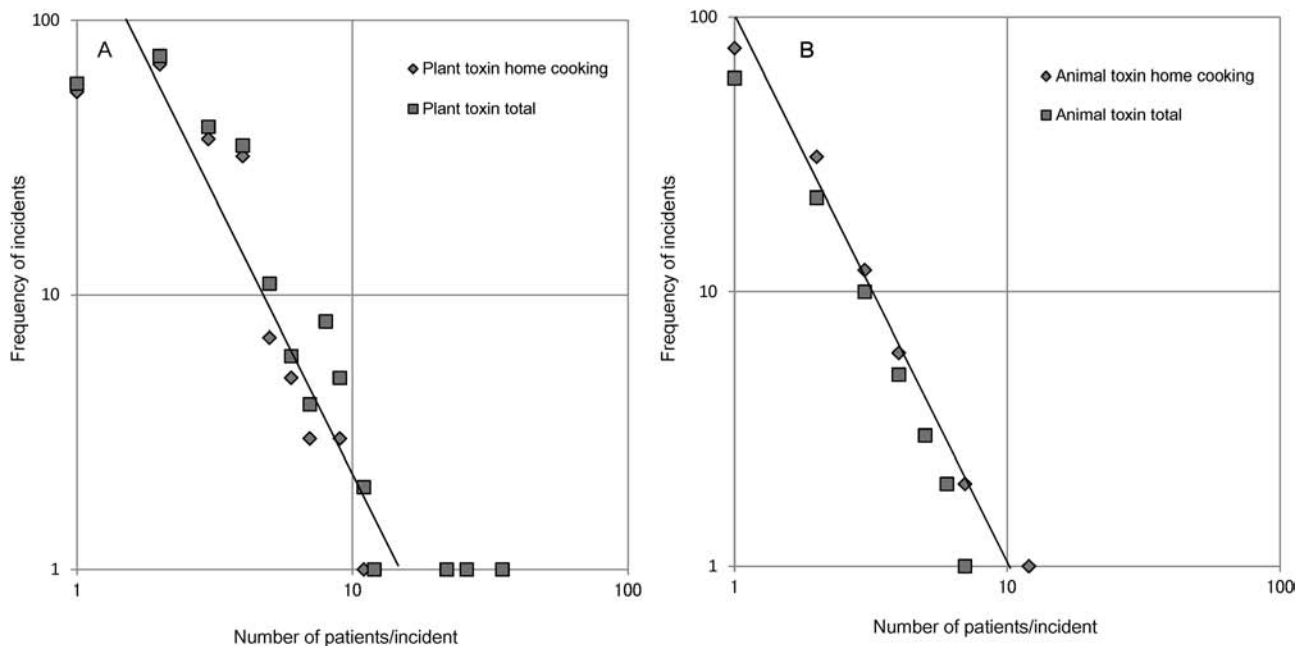
Fig. 3. Log-log plot of the patient number/incident versus its frequency. The number of patients/incident is plotted on the X axis and its frequency on the Y axis. Square symbols represent all the incidents, and diamond symbols represent incidents caused by cooking at home. (A) Food poisoning caused by plant toxins. (B) Food poisoning caused by animal toxins.

ple all natural toxin poisonings have been reported on account of occasional fatality.

The plots for all incidents and home cooking-related incidents almost overlapped, indicating that most cases of food poisoning by natural toxins occurred at home. The curves for both plant toxin and animal toxin food poisonings (panels A and B) were rightward descending straight lines with a slope of 2, i.e., the curves followed the power law. Although the curve for poisoning by plant toxins dipped at X = 1 (one-patient cases), it would be reasonable to consider that the real curve is straight and does not dip, because many cases of food poisoning by plant toxins are generally not as serious as those of food poisoning by animal toxins (e.g., while no patient died from poisoning by plant toxins, three patients died from poisoning by animal toxins during 2008–2010), and many of single-patient cases of plant toxin poisoning may not have been reported.

Foods containing natural toxins are very frequently consumed at home during recreational mushroom hunting or fishing. The size of such events is strongly influenced by the human network. Therefore, it may be natural that the frequency distribution of food poison- ing by natural toxins follows the power law, which is known to rule the human network (3,4).

Conflict of interest None to declare.

REFERENCES

1. Aitken, M., Broadhurst, B. and Hladky, S. (2010): Mathematics for Biological Sciences, Garland Science, NY & London.
2. Codex Alimentarius Commission (2009): Principles and Guidelines for the Conduct of Microbiological Risk Management (MRM), Food Hygiene Basic Texts. 4th ed. World Health Organization/Food and Agriculture Organization.
3. Barabasi, A. (2002): Linked—The New Science of Networks. Perseus Publishing, Cambridge Ma.
4. Watts, D.J. (2003): Six Degrees—The Science of a Connected Age. W.W. Norton & Company, NY & London.