Understanding the statistics of small risks

Compréhension de la statistique des petits risques

by E. Siddall, Atomic Energy of Canada Limited

par E. Siddall, Énergie Atomique du Canada, Limitée

Presented at the Société Française de Radioprotection Conférence on Risk, Avignon, France 1983 October 18-22

1983 October

Présentée lors de la Conférence sur les Risques de la Société Française de Radioprotection, à Avignon, France, du 18 au 22 octobre 1983.

Octobre 1983

Atomic Energy
of Canada Limited
CANDU Operations

L'Énergie Atomique
du Canada, Limitée
Opérations CANDU

Sheridan Park Research Community
Mississauga, Ontario L5K 1B2
Tel. (416) 823-9040

AECL-7541

Understanding the statistics of small risks

Compréhension de la statistique des petits risques

by E. Siddall, Atomic Energy of Canada Limited

par E. Siddall, Énergie Atomique du Canada, Limitée

1983 October

Octobre 1983

Compréhension de la statistique des petits risques

par E. Siddall

Résumé

On se sert d'analyses par la méthode de Monte-Carlo pour montrer les conclusions qu'on peut et ne peut tirer lorsqu'un très petit nombre d'accidents découle d'une dose d'irradiation considérable ou lorsqu'un très petit nombre de personnes allant jusqu'à une seule personne, est exposé à d'autres petits risques. On met en lumière la différence entre l'incertitute relative et l'incertitude absolue. On ne met en jeu aucuns autres principes statistiques.

Octobre 1983

Understanding the statistics of small risks

by E. Siddall

Abstract

Monte Carlo analyses are used to show what inferences can and cannot be drawn when either a very small number of accidents result from a considerable exposure or where a very small number of people, down to a single individual, are exposed to small added risks. The distrinction between relative and absolute uncertainty is illustrated. No new statistical principles are involved.

1983 October

AECL-7541

Atomic Energy of Canada Limited
CANDU Operations

L'Énergie Atomique du Canada, Limitée
Opérations CANDU

Sheridan Park Research Community
Mississauga, Ontario L5K 1B2
Tel. (416) 823-9040

CONTENTS

Page

INTRODUCTION 1

STATISTICAL INFERENCE FROM OBSERVATION 1

Case 1 — An era of nuclear electricity production with reactors of differing risk levels 1

Case 2 — An era of nuclear electricity production with reactors of uniform risk level 2

Comments on Cases 1 and 2

Case 3 — Relative versus absolute uncertainty 2

Comments on Case 3

Case 4 — Added risk to an individual person

Comments on Case 4

CONCLUSION 2

REFERENCES 3

APPENDIX 3

LIST OF TABLES

TABLE 1 NUCLEAR “ERAS” — HIGHER & LOWER RISKS

TABLE 2 NUCLEAR “ERAS” — UNIFORM RISK

TABLE 3 NUMBER OF ACCIDENTS VERSUS RISK LEVEL

TABLE 4

TABLE 5 LIFE SPANS OF 100 INDIVIDUALS — AS CANADA 1979 — WITH AND WITHOUT ADDED RISK

INTRODUCTION

When a large number of accidents (or other stochastic events) occur in each unit of time (or unit of exposure to the probability of the events), it is widely accepted that the numbers can be used as a measure of the risk. For instance, the number of deaths from heart attack (ICDA category 410) amongst 432,900 Canadian males aged 60 to 64 in 1979 was 2356, and the risk would be assessed as 5.44 x 10^{-3} per year within the group at that time.

Many risks, however, including some which are presently given most prominence in our societies, involve very small numbers of accidents and fatalities in each unit of time or exposure, including a high proportion of zeroes. The example considered in the paper is the nuclear electricity industry in the "western" (OECD) world, which has suffered one accident causing appreciable radiation exposure to the public in roughly 24 years, or 1900 reactor-years, of exposure.

A related problem is the risk to an individual arising from some particular source. The number of people involved is one. Considering for the moment the risk to life, which is the dominent problem in safety, that person can only die once. Thus the problem inherently involves two "ones" - ie two very small numbers. The risk from heart attack referred to above may be expressed for the whole population by the numbers shown, but when applied to one individual it becomes a life-or-death proposition. ie 1 or 0.

This paper explores the relationship between the "smooth" risk deduced from large numbers and the "coarse" consequences when these involve small numbers.

STATISTICAL INFERENCE FROM OBSERVATION

A fundamental problem in statistics, and, in fact, in all human actions based on reason, is to decide upon a quantitative general hypothesis from actual limited and uncertain experience. However the process may be disguised, the range of possible future courses of action must ultimately be assessed on the basis of such a general hypothesis so that choices can be made between them.

The difficulty when small numbers are involved either in the experience or the consequences is that the uncertainties become relatively large. There is a human tendency to attempt to circumvent quantitative reasoning in such circumstances. A possible remedy is provided by Monte Carlo simulation, which deals with large and small numbers of stochastic events equally well. It should at least be helpful to human judgment to observe the kind of input models and data which will yield output results similar to those encountered in real-life experience.

The following three cases are examples.

Case 1 — An era of nuclear electricity production with reactors of differing risk levels

The "western" (OECD) nuclear electricity industry grew from the first reactor about 24 years ago to about 200 reactors today for a total of roughly 1900 reactor-years. This exposure is roughly simulated by a model in which 10 new reactors come into service at the beginning of each of 19 discreet years to the present date. By a method based on random numbers, 10 of these 190 reactors are selected to have a risk of accident 3 times greater than the rest. The risks of accident are thus set at .00048 per year for the normal reactors and .00144 for the high risk reactors, these numbers being chosen to fit the observed experience of one accident per 1900 reactor-years or .00053 per reactor year (rounded to 2 digits).

A nuclear "era" is simulated by generating 1900 random numbers (it may be argued that they should be called "pseudo-random") between 0 and 1 from a stated algorithm (see Appendix). Each number is allotted to a stated reactor in a stated year. If the number is less than .0004800 for a normal reactor or .0014400 for a high risk reactor, an accident is deemed to have befallen that reactor in that year.

Table 1 shows the results of 40 such runs, that is, 40 "eras" each of the scale of the one era which has actually occurred. If an accident occurred in an era, the number allotted to the reactor involved is shown in a horizontal position indicating when, in the 19 years of the era, the accident happened. High risk reactors are shown in boxes.

The strange fact that all the accidents to high risk reactors were all grouped only in the low numbered runs is a typical "coincidence" or "chance grouping" which occurs in real life and which is faithfully illustrated by the Monte Carlo method. It will be noted that four reactors, nos. 7, 33, 111 and 113, only one of which was a high risk reactor, appear twice in the Table while 151 reactors do not appear at all. Another coincidence is that in one era, no. 16, reactor no. 1 had an accident in its first year! These striking consequences arose from a known simple pattern of risk and serve as a warning against attaching much significance to particular occurrences or patterns of occurrences.

It will be noted that 41 runs are shown in the Table. This is because a representation of the era which actually happened, complete with one accident, has been inserted in the Table in a random position, disclosed on the last page of the Appendix.

Comparisons between the actual Monte Carlo results and the theoretical statistical expectations from the source data are shown.

Case 2 — An era of nuclear electricity production with reactors of uniform risk level

This was the same as case 1 and used the same series of random numbers. A uniform risk of .00053 per reactor-year was used, that is, an accident was deemed to have occurred if the random number was less than .0005300. The results are shown in Table 2, with a representation of the actual era again inserted.

Comments on Cases 1 and 2

(i) The hypothesis that the experience of the actual era arose from the risk assumptions made in either Case 1 or Case 2 does not in any way conflict with the results obtained.

(ii) The hypothesis that a reactor which had an accident in an era is a high risk reactor as defined in Case 1, or is a reactor in a high-risk subgroup (these criteria are of course identical) is very poorly supported.

(iii) It will be seen that a fixed level of risk can give rise to a considerable variety of accident patterns.

Case 3 — Relative versus absolute uncertainty

With the same model and random number series as in cases 1 and 2, two different uniform risk probabilities were analyzed, ie .001592 and .00032 per reactor-year. (roughly a 5 to 1 ratio, arbitrarily chosen). Table 3 shows the number of accidents per era of 1900 reactor-years in each case. The numbers would be expected to follow a Poisson distribution. The mean and standard deviation are calculated for each risk level and are compared with the theoretically expected values.

Comments on Case 3

It will be seen that the standard deviation is a higher fraction of the mean in the case of the lower risk. This illustrates "statistical coarseness" and is of course the reason why small numbers tend to be distrusted.

However, it will also be seen that the absolute value of the standard deviation diminishes as the risk diminishes, in the Monte Carlo study as in the theoretical expectation. It is surely the absolute uncertainty which is of importance in practice.

Case 4 · Added risk to an individual person

The basic total risk is taken to depend on age as shown in the second and third columns of Table 4, which are simplified from the Candian life table for 1979. As developed in ref 3, safety in a society can in practice be measured by the probability of survival to age 65, so that mortality beyond that age is not considered in this case. The added risk is considered to be uniform and independent of age. It is chosen to increase the risk of premature death (ie before age 65) by 10% compared with the normal case. The consequential risk pattern is shown in the fourth and fifth columns.

For each 100 individuals, a random number between 0 and 1 was allotted (see Appendix). The magnitude of the number was compared with the survival functions of the third and fifth columns to obtain a length of life for each individual, zero representing zero life and .2070203 (= 1-.7929797), for example, representing survival to age 65.0 in the normal population. Table 5 shows the results. "A" signifies survival beyond age 65. Two of the columns (50 people) apply to individuals in the normal population and two columns apply to individuals exposed to the higher risk. The reader should try to determine which individuals and which columns (groups of 25 individuals) arose from which risk levels. The correct answer is given in the last page of the Appendix.

Comment on Case 4

Concentrating attention on any one individual in Table 5 will at once make it clear that there is no possibility whatever of deciding whether or not he was exposed to the added 10% of risk. It is not even possible to determine which column of 25 was exposed to which risk level.

It is therefore clear that a 10% added risk of premature death makes no perceptible difference to the life span of an individual; the effect is fundamentally and completely masked by the "statistical coarseness" of the ordinary risks of living.

CONCLUSION

Monte Carlo simulation offers the possibility of an important contribution to logic, clear thinking and sense of perspective in the science of human safety.

As understanding of risk and safety slowly accumulates, it becomes clear that all risks and all safety benefits tend to be unequally distributed, as shown, for instance, in ref. 3. Greater safety for the whole society must sometimes involve an increase in a component of risk to some individuals. This paper shows that we must think very carefully before we allow excessive concern for individual risk to impede measures such as wealth production which benefit safety in the society as a whole.

REFERENCES

(1) **Naylor, T.H.** et al. *Computer simulation techniques* John Wiley & Sons, New York, 1966.

(2) **Abramovitch, N.** and **Stegun, I.A.** (eds) *Handbook of mathematical functions* Dover Publications Inc. New York 1965.

(3) **Siddall, E.** *Safety policy in the production of electricity.* Proc. International meeting on thermal reactor safety, Chicago, ANS-CNS-ENS-JAES, August 29 - September 2 1982.

APPENDIX

It is intended that the random processes used in this paper should be reproducible, (rigorously, of course, this means that they cannot be random!). The random number algorithm is of the mixed linear congruential form discussed in ref 1., as follows.

$n_{x+1} = [an_x + c]$ modulo 1

n is a 10 digit number less that 1 and equal to or greater than 0.

a is chosen to be 981 which is the highest number = 1 modulo 20 which keeps the calculation within the 13 digit capacity of the machine used.

Other constants were taken from Table 26.11 (of random numbers) from ref. 2, as follows:-

c = .2668774223, being groups 1 and 2

For Cases 1 and 2, n_0 (seed) = .4354645699, which are groups 3 and 4.

For Case 3, n_0 (seed) = .9446982125, which are groups 5 and 6.

In case 1 the 10 high risk reactors were selected from the total of 19 by taking triples as shown from pairs of numbers in the second block of Table 26.11 of ref. 2, taking the first digit as zero if even and 1 if odd, as follows:-

846	865	763	6	46.	65.	163.
323	261	986	7	123.	61.	186
713	454	200	2	113.	54.	
969	978	437	9	169.	178.	37
279	912	145	9	79	112	145.

In Tables 1 and 2, Three Mile Island unit 2 is given the number 141 and the actual era is line 36.

In Table 5, the first two columns apply to members of the normal population and the third and fourth columns to those with added risk.

1	
2	$\boxed{113}$
3	$\boxed{113}$ $\boxed{169}$
4	
5	$\boxed{61}$
6	
7	40
8	31 $\boxed{123}$
9	
10	28
11	33
12	
13	45
14	12
15	48 76
16	1 41
17	
18	151
19	67 40
20	
21	20 135
22	35
23	19 182
24	81
25	
26	111
27	6 31
28	
29	39 75 92
30	
31	124
32	24 33 122
33	16 111
34	5
35	—
36	141
37	—
38	12
39	7
40	71,5
41	7

NO. IN ERA	ACTUAL	EXPECTED
0	12	14.61
1	16	14.71
2	10	7.41
3	2	2.49

TOTALS	ACTUAL	EXPECTED
HIGH RISKS	5	5.75
NORMAL	37	34.53
TOTAL	42	40.28

NOTES:

1. EACH SERIALLY NUMBERED LINE IS AN ERA OF 190 REACTORS, COMPLETED 10 AT THE BEGINNING OF EACH YEAR FOR 19 YEARS, AMOUNTING TO 1900 REACTOR-YEARS.
2. THE POSITION OF A NUMBER SHOWS APPROXIMATELY WHEN, IN THE 19 YEAR ERA, AN ACCIDENT OCCURRED. THE NUMBER INDICATES WHICH REACTOR WAS INVOLVED.
3. $\boxed{}$ INDICATES A REACTOR WITH 3 X THE GENERAL LEVEL OF RISK.

TABLE 1 NUCLEAR "ERAS" — HIGHER & LOWER RISKS

1	
2	113
3	
4	
5	
6	
7	40
8	31
9	
10	28
11	33
12	
13	45
14	12
15	48 76
16	41
17	
18	151
19	67
20	
21	20 135
22	35
23	19 182
24	81
25	128
26	111
27	6 26 31
28	
29	39 75 92
30	
31	124
32	24 33 122
33	16 111
34	5 74
35	
36	141
37	122
38	12
39	7
40	71,5
41	7

NO. IN ERA	ACTUAL	EXPECTED
0	12	14.61
1	18	14.71
2	7	7.41
3	3	2.49

TOTAL	
ACTUAL	EXPECTED
41	40.28

NOTES:

1. EACH SERIALLY NUMBERED LINE IS AN ERA OF 190 REACTORS, COMPLETED 10 AT THE BEGINNING OF EACH YEAR FOR 19 YEARS, AMOUNTING TO 1900 REACTOR-YEARS.
2. THE POSITION OF A NUMBER SHOWS APPROXIMATELY WHEN, IN THE 19 YEAR ERA, AN ACCIDENT OCCURRED. THE NUMBER INDICATES WHICH REACTOR WAS INVOLVED.
3. ▭ INDICATES A REACTOR WITH 3 X THE GENERAL LEVEL OF RISK.

TABLE 2 NUCLEAR "ERAS" — UNIFORM RISK

ERA NO.	RISK LEVEL		ERA NO.	RISK LEVEL	
	.00159	.00032		.00159	.00032
1	1	0	21	5	0
2	4	0	22	3	0
3	2	0	23	4	0
4	2	0	24	2	0
5	3	0	25	3	0
6	3	0	26	2	1
7	4	1	27	3	2
8	6	1	28	0	0
9	3	0	29	10	2
10	5	0	30	0	0
11	4	1	31	3	1
12	3	0	32	7	2
13	4	1	33	4	1
14	5	0	34	4	1
15	2	2	35	1	0
16	3	2	36	2	0
17	0	0	37	1	1
18	1	1	38	4	1
19	3	2	39	4	1
20	2	0	40	4	0

		ACTUAL	EXPECTED
RISK LEVEL .00159	MEAN STD. DEV.	3.1500 1.9289 61.23 %	3.0252 1.7393 57.49 %
RISK LEVEL .00032	MEAN STD. DEV.	.6000 .7442 124.03 %	.6080 .7797 128.24 %

TABLE 3 NUMBER OF ACCIDENTS VERSUS RISK LEVEL

AGE	NORMAL RISK		AGE	ADDED RISK	
	RISK PER YEAR	SURVIVAL (LIFETABLE)		RISK PER YEAR	SURVIVAL (LIFETABLE)
0		1.0000000	0		1.0000000
	.010909			.0113147	
1		.9891502	1		.9887489
	.000580			.0009857	
5		.9868580	5		.9848580
	.000342			.0007477	
20		.9818084	20		.9738732
	.001215			.0016207	
40		.9582368	40		.9428087
	.003182			.0035877	
50		.9282233	50		.9095792
	.006469			.0068747	
55		.8986777	55		.8788419
	.009724			.0101297	
60		.8560210	60		.8354291
	.015292			.0156977	
65		.7929797	65		.7723335

NOTES:

1. NORMAL RISK IS APPROXIMATELY THAT OF CANADA IN 1979.
2. THE ADDED RISK IS .0004057 PER YEAR WHICH INCREASES THE RISK OF PREMATURE (BEFORE AGE 65) DEATH BY 10%.

TABLE 4

1	12.967	26	A	51	A	76	A
2	A	27	A	52	A	77	A
3	A	28	A	53	A	78	A
4	A	29	60.345	54	A	79	A
5	A	30	A	55	61.915	80	A
6	A	31	A	56	A	81	A
7	A	32	A	57	A	82	A
8	A	33	41.557	58	A	83	17.418
9	58.411	34	A	59	A	84	30.280
10	A	35	55.526	60	58.108	85	A
11	57.697	36	47.673	61	A	86	A
12	A	37	A	62	A	87	61.152
13	A	38	A	63	59.626	88	A
14	58.410	39	39.827	64	A	89	A
15	A	40	35.932	65	A	90	64.623
16	A	41	A	66	A	91	A
17	64.457	42	A	67	A	92	A
18	A	43	A	68	A	93	A
19	21.747	44	A	69	13.591	94	A
20	58.737	45	55.046	70	A	95	40.354
21	A	46	43.400	71	A	96	A
22	25.311	47	A	72	A	97	64.955
23	A	48	A	73	A	98	32.506
24	A	49	A	74	29.575	99	A
25	.783	50	A	75	A	100	29.818

'A' signifies 65 years or longer

TABLE 5 LIFE SPANS OF 100 INDIVIDUALS — AS CANADA 1979 — WITH AND WITHOUT ADDED RISKS

The International Standard Serial Number

ISSN 0067—0367

has been assigned to this series of reports.

To identify individual documents in the series
we have assigned an AECL—number.

Please refer to the AECL—number when
requesting additional copies of this document
from

Scientific Document Distribution Office
Atomic Energy of Canada Limited
Chalk River, Ontario, Canada
K0J 1J0

Price $2.00 per copy

Le numéro de série standard international

ISSN 0067—0367

a été attribué à cette série de rapports.

Pour identifier les rapports individuels faisant partie
de cette série nous leur avons attribué un numéro AECL-

Veuillez faire mention du numéro AECL- si vous
demandez d'autres exemplaires de ce rapport
au

Service de Distribution des Documents Officiels
L'Énergie Atomique du Canada, Limitée
Chalk River, Ontario, Canada
K0J 1J0

Prix $2.00 par exemplaire