Brit. J. prev. soc. Med. (1970), **24**, 163-168

'CLUSTERING' OF MULTIPLE SCLEROSIS CASES BY DATE AND PLACE OF BIRTH

G. AMARH ASHITEY, M.D. (Belfast), D.Obst.R.C.O.G., D.P.H.
Research Fellow

AND

GILBERT MACKENZIE
Experimental Officer
Department of Social and Preventive Medicine, The Queen's University of Belfast

IT has been suggested that a 'slow virus' may cause multiple sclerosis, and experiments on the possibility of transmitting the agent are continuing (Campbell, Norman, and Sandry, 1963; Dick, McAlister, McKeown, and Campbell, 1965; Field, 1966, 1968, 1969). Further, some circumstantial evidence suggests that the disease may be acquired in infancy even though the clinical manifestations appear much later (Millar, 1966; Dean, 1967; Kurtzke, 1968).

These two suggestions could be combined into a single, simple, and testable hypothesis which, if true, should logically result in the 'clustering' (i.e. groups of cases closely associated in time and place) of cases by date and place of birth simultaneously. This paper reports the results of an investigation of the temporal-spatial distribution of multiple sclerosis in Northern Ireland. The statistical techniques used are basically those described by Ederer, Myers and Mantel (1964) and by David and Barton (1966).

DATA

ASCERTAINMENT

Since 1948, efforts have been made to ascertain all cases of multiple sclerosis in Northern Ireland and a central register of all the ascertained cases (living and dead) has been compiled. On 1 October 1968, there were 1,418 cases on the register. Most of them were ascertained through a countrywide prevalence survey in 1949–53 (Allison and Millar, 1954). The remainder have been ascertained since 1954, mainly by these two consultant neurologists, through general practitioner referrals, consultant clinics, hospital records, and post-mortem reports. In the absence of any specific diagnostic test it is not possible to estimate the degree of ascertainment that has been achieved. However, the structure and administrative arrangements of the Health Service in Northern Ireland are such that, using the combination of methods listed above and the diagnostic criteria of Allison and Millar (which were adopted by the World Federation of Neurologists for epidemiological studies (Allison, 1960)), very few cases are likely to have been missed.

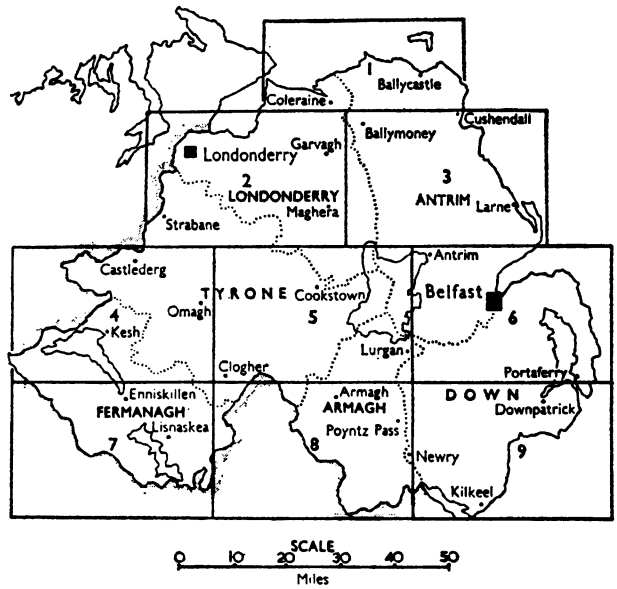
The place and year of birth of patients are among the data routinely obtained in the course of ascertainment. These were extracted from the patients' records, and in the few cases where they had not been recorded, the information was obtained by means of a postal survey conducted among the surviving cases.

MULTIPLE SCLEROSIS CASES STUDIED

From the total of 1,418 cases, two groups, A and B, were selected for this investigation.

Group A comprised all 783 patients born within the area now known as Northern Ireland during the years 1901–25. This period was chosen as multiple sclerosis manifests itself mainly between the ages of 20 and 40 years (Brain, 1966; Acheson, 1965). Consequently, many individuals born after 1925 and destined to develop the disease would not have manifested it at the time of this investigation. In addition some multiple sclerosis patients born before 1901 would have died before 1948 (the year in which ascertainment began) and were therefore missed. Finally, the data necessary for the valid application of the method of Ederer *et al.* (1964) were available on all group A cases.

For the application of the method of David and Barton (1966) a more circumscribed group of cases with more detailed information is desirable. Accordingly, a sub-group B, of group A, comprising 109 multiple sclerosis patients born in Northern Ireland during 1921–25 whose exact day of birth could be traced, was selected.



Geographical areas
1 North Coast
2 Londonderry
3 Mid Antrim
4 Omagh
5 East Tyrone
6 Lagan Valley
7 Enniskillen
8 Armagh
9 South Down
10 Belfast
.....Administrative boundaries

METHODS

The two statistical techniques employed have been used widely in recent years to study the space-time distributions of relatively rare diseases, particularly leukaemia (Ederer *et al.*, 1964; Ederer, Myers, Eisenberg, and Campbell 1965; Barton, David, and Merrington, 1965; Lock and Merrington, 1967; Merrington and Spicer, 1969). They require, therefore, no detailed description here and only their application to the present data is presented.

METHOD OF EDERER *et al.* (1964)

The method of Ederer *et al.* (1964) requires *a priori* the definition of the space-time units to be used. In general, a space-time unit will consist of a suitably defined area studied for a period of time. The most suitable space-time units derivable from the data available were the administrative areas and geographical areas shown in the Figure, each studied for the period of one calendar year. Both administrative and geographical units were used because morbidity and mortality data are normally handled on an administrative basis, although in physical terms the spread of infectious diseases is more likely to follow the geographical contours of the country. Administratively, Northern Ireland was considered to consist of seven rather than eight areas—Londonderry County Borough and Londonderry County were studied as one area. The geographical areas shown in the Figure are based on the Ordnance Survey map of Northern Ireland, the tenth area is Belfast. The time interval of one calendar year was selected because, for most of the cases, only the year of birth was known.

To eliminate the effects of possible time trends in the number of cases born in a particular area and for computational convenience the period 1901–25 was divided up into five non-overlapping quinquennial periods, 1901–05, 1906–10, 1911–15, 1916–20 and 1921–25. From these divisions $5 \times 7 = 35$, five-year administrative units* and $5 \times 10 = 50$, five-year geographical units were obtained. Thus, each of these larger units comprised five of the appropriate calendar year space-time units defined above. The relevant data together with K, the total number of cases occurring in any five-year unit, are presented in Tables I and II.

The statistics m_1 and m_2, defined respectively as the observed maximum number of cases occurring in any one year of a five-year unit and the observed maximum number of cases occurring in any two consecutive years of a five-year unit, were employed as indices of clustering. For example, the values of m_1, m_2, and K for the quinquennium 1901–05 in area 10 are 12, 20, and 42 respectively (Table II).

The null hypothesis that the numbers of cases occurring in each five-year unit were randomly distributed was tested. The test statistic proposed by Ederer *et al.* (1964) is a continuity corrected x^2 with one degree of freedom, defined as,

$$x^2 = \frac{(|\Sigma m_1 - \Sigma \mu_1| - 0)^2}{\Sigma \sigma_1} \qquad (1)$$

*Where there is no ambiguity, space-time unit has been abbreviated to unit.

TABLE I

DISTRIBUTION OF 783 ASCERTAINED MULTIPLE SCLEROSIS CASES BORN IN NORTHERN IRELAND 1901–25 (GROUP A) BY ADMINISTRATIVE AREA OF BIRTH

(Quinquennial totals, K, in brackets)

Year of Birth	Administrative Area of Birth (see map)							Total
	BE	AN	AR	DO	FE	LO	TY	
1901	7	2	1	4	0	1	3	18
1902	12	5	2	8	2	1	2	32
1903	8	9	3	7	1	4	1	33
1904	7	4	5	10	5	1	4	36
1905	8	5	5	5	5	0	2	30
	(42)	(25)	(16)	(34)	(8)	(9)	(15)	(149)
1906	8	6	3	4	1	4	4	30
1907	8	6	2	4	1	5	3	29
1908	12	7	3	11	3	3	2	41
1909	6	8	5	5	0	4	8	36
1910	9	7	4	10	3	2	5	40
	(43)	(34)	(17)	(34)	(8)	(18)	(22)	(176)
1911	9	8	5	4	5	1	0	32
1912	8	5	3	9	0	1	2	28
1913	4	9	6	8	2	1	4	34
1914	10	5	2	8	4	2	4	35
1915	4	4	5	6	2	3	5	29
	(35)	(31)	(21)	(35)	(13)	(8)	(15)	(158)
1916	4	8	4	6	1	4	8	35
1917	5	2	1	6	3	2	4	23
1918	7	4	2	7	4	2	1	27
1919	9	3	5	2	1	5	5	30
1920	5	7	4	12	1	4	5	38
	(30)	(24)	(16)	(33)	(10)	(17)	(23)	(153)
1921	11	3	1	5	1	2	12	35
1922	5	7	7	2	1	6	5	33
1923	8	3	6	5	0	3	4	29
1924	6	3	3	1	4	2	4	23
1925	3	3	3	1	3	5	9	27
	(33)	(19)	(20)	(14)	(9)	(18)	(34)	(147)

BE = Belfast, AN = Antrim, AR = Armagh, DO = Down, FE = Fermanagh, LO = Londonderry, and TY = Tyrone

TABLE II

DISTRIBUTION OF 783 ASCERTAINED MULTIPLE SCLEROSIS CASES BORN IN NORTHERN IRELAND 1901–25 (GROUP A) BY GEOGRAPHICAL DIVISION OF BIRTH

(Quinquennial totals, K, in brackets)

Year of Birth	Geographical Area of Birth (see map)										Total
	1	2	3	4	5	6	7	8	9	10	
1901	0	4	2	0	0	3	0	1	1	7	18
1902	1	0	2	0	3	4	2	3	5	12	32
1903	1	3	7	0	2	5	2	4	1	8	33
1904	1	2	2	1	2	6	4	5	6	7	36
1905	0	3	4	1	7	4	0	1	2	8	30
	(3)	(12)	(17)	(2)	(14)	(22)	(8)	(14)	(15)	(42)	(149)
1906	1	5	4	0	5	3	1	1	2	8	30
1907	1	5	4	1	2	5	1	1	1	8	29
1908	0	3	4	1	4	8	2	2	5	12	41
1909	1	4	4	4	5	4	0	5	3	6	36
1910	0	2	6	1	6	6	3	2	5	9	40
	(3)	(19)	(22)	(7)	(22)	(26)	(7)	(11)	(16)	(43)	(176)
1911	1	1	4	0	1	7	2	7	0	9	32
1912	2	1	3	0	2	5	0	4	3	8	28
1913	2	1	3	1	5	7	1	6	4	4	34
1914	0	2	1	2	4	7	3	3	3	10	35
1915	1	3	3	1	5	5	2	4	1	4	29
	(6)	(8)	(14)	(4)	(17)	(31)	(8)	(24)	(11)	(35)	(158)
1916	2	5	4	3	7	6	0	3	1	4	35
1917	0	1	3	0	4	2	3	2	3	5	23
1918	2	2	1	4	1	5	0	1	4	7	27
1919	2	2	1	1	7	4	1	2	1	9	30
1920	0	4	3	4	3	9	0	6	4	5	38
	(6)	(14)	(12)	(12)	(22)	(26)	(4)	(14)	(13)	(30)	(153)
1921	1	2	1	6	6	6	1	1	0	11	35
1922	2	4	4	2	7	5	0	4	0	5	33
1923	1	4	2	1	6	2	0	5	0	8	29
1924	0	1	1	2	3	4	4	2	0	6	23
1925	2	4	0	6	4	3	1	2	0	5	27
	(6)	(15)	(8)	(17)	(26)	(20)	(6)	(14)	(0)	(35)	(147)

where, for a particular five-year unit, μ_1 is the expected value (hereafter denoted by E) of m_1 given K and σ_1^2 is the variance (hereafter denoted by VAR) of m_1 given K, the summation extending over all admissible five-year units. The parameters μ_1 and σ_1^2 are obtained from the conditional probability distribution of m_1 given K by a method credited to Feller (1957) and detailed in the appendix. Although equation (1) may be used as a test criterion for m_2 by substituting m_1, μ_1, and σ_1^2 with m_2, μ_2 and σ_2^2, these latter parameters cannot be obtained directly by Feller's method.

As K becomes large the generation of the exact occupancy distribution of each five-year unit on the computer* becomes lengthy. However, the computing time may be substantially reduced by estimating the parameters μ_1, σ_1^2, μ_2, and σ_2^2 using quasi-occupancy distributions generated by Monte Carlo methods (Naylor, Balintfy, Burdick, and Chu, 1966). This approach was adopted for the estimation of the parameters $E(m_1/K)$ and $VAR(m_1/K)$ for $K>35$, in addition to $E(m_2/K)$ and $VAR(m_2/K)$ for all K. As the estimation procedure involved 1,000 simulated trials, the degree of approximation was small and the reliability of the test criterion was not affected.

Using this technique, Ederer *et al.* (1965) have demonstrated the clustering of cases of poliomyelitis and infectious hepatitis in Connecticut.

*I.C.L. 1907, Q.U.B.

METHOD OF DAVID AND BARTON (1966)

The David and Barton (1966) space-time interaction test, which was used to analyse group B cases, treats the space co-ordinates as a randomization set and postulates that all possible allocations of a given time to a given space point are equally likely. The statistical test criterion used is the ratio of the average squared distance between space points within time clusters to the overall average. This method, therefore, unlike that proposed by Knox (1963), does not require a definition of a unit of distance but does require divisions of time to be made prior to the analysis. Where the incubation period of a particular disease is known, the interval of time chosen may be logically based on such knowledge, but where such a period is unknown, as it is here, the selection must be arbitrary.

TABLE III

DISTRIBUTION OF 109 ASCERTAINED MULTIPLE SCLEROSIS CASES BORN IN NORTHERN IRELAND, 1921–25 (GROUP B)

Year of Birth	Cases Traced and Analysed	Cases Not Traced	All Ascertained Cases
1921	28	7	35
1922	25	8	33
1923	19	10	39
1924	17	6	23
1925	20	7	27
Total	109 (74%)	38 (26%)	147 (100%)

The exact day of birth, required for this test, was successfully obtained in 109 (74%) of the 147 cases born during 1921–25 (Table III). The space co-ordinates (x, y) of the place of birth of these cases were taken from the Ordnance Survey map, 3rd series (Ashitey, 1969). These were arranged consecutively in time order and divided into time groups using each of the time intervals, 8 days, 18 days, 31 days, 41 days, 51 days, and 61 days. The test criterion Q was calculated for each time interval using a computer programme kindly supplied to us by Dr. Maxine Merrington.

This test criterion has previously been used to show clustering of cases of measles and of poliomyelitis but not of acute leukaemia (Barton *et al.*, 1965).

RESULTS

GROUP A CASES

Pertinent results are summarized in Table IV. No significant results were obtained, the calculated values of x^2 for both administrative and geographical 5-year units being less than the usually acceptable value (3·84) for one degree of freedom. The generally small deviation from expectation obtained, particularly for geographical divisions, may be indicative of a mechanism leading to a quinquennium to quinquennium uniformity in the maximum number of people in one year who were born in those areas and subsequently developed the disease. It was therefore concluded, on the evidence afforded by these data, that multiple sclerosis cases in Northern Ireland were not clustered, by their year and place of birth simultaneously, during the period investigated.

TABLE V

RESULTS OF ANALYSIS OF GROUP B CASES USING METHOD OF DAVID AND BARTON (1966)

Interval between Time Clusters (days)	No. of Clusters	Q	d^*
8	70	0·79	−1·70
18	42	0·93	−0·96
31	13	0·97	−0·61
41	7	0·98	−0·83
51	5	0·99	−0·31
61	4	0·99	−0·73

* $d = (Q - 1)/S.E.$ of Q

GROUP B CASES

Table V shows the values of the test criterion Q obtained for the time intervals studied. Assuming that Q is approximately normally distributed, d^* may be referred to the normal probability scale. Since none of the d values is less than −2 (David and Barton, 1966), it was concluded that the null hypothesis cannot be rejected, i.e., there is no significant evidence of clustering of the multiple sclerosis cases studied by this method.

DISCUSSION AND CONCLUSION

The results of the present investigation provided no evidence for the rejection of the null hypothesis that multiple sclerosis cases in Northern Ireland were randomly distributed by date and place of birth simultaneously. It should be emphasized that this finding must be viewed against the assumption that the disease is acquired in infancy as well as the statistical techniques employed.

With the method of Ederer *et al.* (1964), the space-time units chosen are an integral part of the hypothesis tested, and consequently the results obtained would not necessarily be valid for units defined differently from those employed here. While the degree of completeness of ascertainment of cases would influence the results, the data used were thought to be reliable.

* $d = (Q-1)/S.E.$ of Q

TABLE IV

RESULTS OF ANALYSIS OF GROUP A CASES BY METHOD OF EDERER *et al.* (1964)

Area	No. of 5-year Units	No. of 5-year Units with Two or More Cases	Analysis of Index m_1				Analysis of Index m_2			
			Σm_1	$E(\Sigma m_1)$	x^2	$P(x^2)$	Σm_2	$E(\Sigma m_2)$	x^2	$P(x^2)$
Administrative	35	35	252	244·15	1·090	0·3 > p > 0·2	400	401·94	0·33	0·9 > p > 0·8
Geographical	50	49*	254	254·23	0·001	0·98 > p > 0·95	411	415·99	0·27	0·7 > p > 0·5

*One geographical 5-year unit (Table II) had zero cases and was omitted from the analysis.

The method of David and Barton (1966) has the property of testing interaction whilst being substantially unaffected by irregularities of the separate time and space distributions. This, of course, is an advantage where ascertainment is incomplete or where cases are geographically clustered in towns and villages. In this investigation 109 of the 147 known cases born between 1921 and 1925 could be included in the analysis. However, there appears to be no obvious reason why these omissions should obscure rather than magnify any clustering that might have occurred.

Although the demonstration of clustering of cases would have provided an important piece of evidence in support of the hypothesis that multiple sclerosis is due to a 'slow virus' infection acquired in infancy, other factors including infectious bacteria, radiation, poisons, etc., may also give rise to clusters of cases. Similarly, the present negative result is not conclusive, by itself, in excluding a viral aetiology, since if the 'virus' is endemic and common, then the people affected by it might be randomly distributed. This latter possibility is unlikely to be true of multiple sclerosis which is relatively rare; but matters may be complicated further by postulating a causative virus which attacks many people, with only a fraction of them developing a neurological complication, as in the case of measles. However, the absence of clustering of cases by date and place of birth simultaneously is interesting because it supports other epidemiological work (Ashitey and Millar, 1970) which suggests that if the disease is acquired in infancy or even in adolescence, then the aetiological factor does not exhibit 'epidemic' characteristics.

SUMMARY

An investigation was carried out into the time-space clustering of cases of multiple sclerosis, using the date and place of birth of affected persons. Altogether 783 ascertained cases born in Northern Ireland during the period 1901–25 were studied by two statistical techniques. No clustering of cases was detected. The significance of the results is discussed.

We thank Dr. J. H. D. Millar and Dr. R. S. Allison for allowing us to study their records; Dr. Maxine Merrington, Department of Statistics, University College, London, for the computer programme for the David and Barton (1966) test; the computing laboratory of Queen's University, Belfast, for facilities, and Mrs. Eileen MacKenzie for secretarial assistance. Finally, we are grateful to Professor Peter Froggatt for his advice and helpful criticism during the study and in the preparation of this paper.

REFERENCES

ACHESON, E. D. (1965). The epidemiology of multiple sclerosis. In *Multiple Sclerosis: A Reappraisal*, p. 3. Edited by McAlpine, D., Lumsden, C. E., and Acheson, E. D. Livingstone, Edinburgh and London.

ALLISON, R. S. (1960). In *Report on the Geomedical conference in Copenhagen* 1959. Edited by Hullested, K. *Acta Psychiat. scand.*, 35, Suppl. 147, p. 18.

——, and MILLAR, J. H. D. (1954). Prevalence and familial incidence of disseminated sclerosis. A report to the Northern Ireland Hospitals Authority on the results of a three-year survey. *Ulster med. J.*, **23**, Suppl. 2.

ASHITEY, G. A. (1969). *Some epidemiological characteristics of multiple sclerosis in Northern Ireland.* An M.D. Thesis, The Queen's University of Belfast.

——, and MILLAR, J. H. D. (1970). Multiple sclerosis in Northern Ireland. A study of the date and place of birth of patients. *Ulster med. J.*, **39**, 55.

BARTON, D. E., DAVID, F. N., and MERRINGTON, M. (1965). A criterion for testing contagion in time and space. *Ann. hum. Genet.*, **29**, 97.

BRAIN, W. R. (1966). *Diseases of the Nervous System*, 6th ed. London University Press, London.

CAMPBELL, A. M. G., NORMAN, R. M., and SANDRY, R. J. (1963). Subacute encephalitis in an adult associated with necrotising myelitis and results of animal innoculation experiments. *J. Neurol. Neurosurg. Psychiat.*, **26**, 439.

DAVID, F. N., and BARTON, D. E. (1966). Two space-time interaction tests for epidemicity. *Brit. J. prev. soc. Med.*, **20**, 44.

DEAN, G. (1967). Annual incidence, prevalence, and mortality of multiple sclerosis in white South-African born and in white immigrants to South Africa. *Brit. med. J.*, **2**, 724.

DICK, G., MCALISTER, J. J., MCKEOWN, F., and CAMPBELL, A. M. G. (1965). Multiple sclerosis and scrapie. *J. Neurol. Neurosurg. Pshchiat.*, **28**, 560.

EDERER, F., MYERS, M. H., and MANTEL, N. (1964). A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics*, **20**, 626.

——, ——, EISENBERG, H., and CAMPBELL, P. C. (1965). Temporal-spatial distribution of leukemia and lymphoma in Connecticut. *J. nat. Cancer Inst.*, **35**, 625.

FELLER, W. (1957). *An Introduction to Probability Theory and its Applications, Vol.* 1, 2nd ed. Wiley, New York.

FIELD, E. J. (1966). Transmission experiments with multiple sclerosis: an interim report. *Brit. med. J.*, **2**, 564.

—— (1968). Multiple sclerosis-virological considerations, together with a note on the passage of Kuru. *Proc. roy. Soc. Med.*, **61**, 940.

—— (1969). Slow virus infection of the nervous system. *Med. News (N.Y.)*, March 21, 1969.

KNOX, G. (1963). Detection of low intensity epidemicity. Application to cleft lip and palate. *Brit. J. prev. soc. Med.*, **17**, 121.

KURTZKE, J. F. (1968). Multiple sclerosis and infection from an epidemiologic aspect. *Neurology (Minneap.)*, **18**, Suppl., p. 170.

LOCK, S. P., and MERRINGTON, M. (1967). Leukaemia in Lewisham (1957-63). *Brit. med. J.*, **3**, 759.

MERRINGTON, M., and SPICER, C. C. (1969). Acute leukaemia in New England: an investigation into the clustering of cases in time and place. *Brit. J. prev. soc. Med.*, **23**, 124.

MILLAR, J. H. D. (1966). Multiple sclerosis, two high risk areas in Northern Ireland. *J. Irish med. Ass.*, **59**, 138.

NAYLOR, T. H., BALINTFY, J. L., BURDICK, D. S., and CHU, K. (1966). *Computer Simulation Techniques*. Wiley, New York.

APPENDIX

The following method due to Feller (1957) was used in the present investigation to obtain the conditional probability distribution m_1, when K, the total number of cases occurring in any five-year unit, was less than or equal to 35. For values of K beyond 35, recourse to Monte Carlo methods is necessary in order to avoid excess processor time.

Feller's formula for n (=5) years, and where the cells are arranged in descending order of occupying cases $k_1, k_2, \ldots\ldots, k_5$, is

$$P(k_1, k_2, \ldots\ldots, k_5) = 5! \times K! \times (\pi_i k_i!)^{-1} \times (\pi_j j!)^{-1} \times 5^{-K}$$

where n_j ($j = 0, \ldots\ldots, K$) is the number of years containing exactly j cases, and k ($i = 1, \ldots\ldots, 5$) is the i^{th} occupancy number.

Since the occupancy numbers have been arranged in descending order of occupying cases, m_1 is identically equal to k_1. Thus, denoting the conditional probability distribution of m_1 given K by $P(m_1|K)$, the first two moments about the mean of this distribution, μ_1 and σ_1^2, may be computed by definition, i.e.,

$$\mu_1 = \Sigma m_1 . P(m_1|K)$$

$$\sigma_1^2 = \Sigma m_1^2 . P(m_1|K) - \mu_1^2$$

where in each case the summation extends over the admissible range of m_1 values, such that $\Sigma P(m_1|K) = 1$.