A STATISTICAL ANALYSIS OF AGE DISTRIBUTION OF ASTHMATICS.

By W. F. HARVEY, W. O. KERMACK, D. M. LYON and A. G. McKENDRICK.

(With 5 Figures.)

During the last eight years a group of workers has been collecting information regarding asthma in the Edinburgh area. The work has been directed by a committee consisting of Prof. D. Murray Lyon, Lieut.-Colonel A. G. McKendrick, I.M.S., Lieut.-Colonel W. F. Harvey, I.M.S., Dr J. S. Fraser, Dr R. Cranston Low, Prof. J. Lorrain Smith, and Prof. J. C. Meakins, and the present communication is an analysis of the statistical data collected referring to age at primary onset, and age at time of examination.

The figures refer to a group of asthmatics reporting at the Royal Infirmary of Edinburgh during the period.

Asthma cannot be considered to be a simple disease entity, but is a condition characterised by a symptom complex, and is dependent in different individuals upon various disease mechanisms. The cases included have been diagnosed from a history of recurring attacks of paroxysmal dyspnoea, and include examples of so-called ideopathic asthma and of bronchitic asthma. Patients suffering from dyspnoea obviously due to cardiac or renal disorders have been excluded.

We are not prepared to deal with the question as to whether these figures may be taken as typical of the population of the Edinburgh district as a whole. The results must be taken as referring to the community from which the data have been drawn. It does not seem possible to decide whether that community consists of the whole or only one very selected part of the population of the Edinburgh area. In one respect the figures are admittedly somewhat defective. The number of patients under 10 years of age is smaller than might be expected, since many young children in this area are dealt with in a special hospital. It will be seen that this deficiency comes to light in the following analysis.

The figures which have been obtained are summarised in Tables I and II which refer to males and females respectively.

In each table is given the number of patients who presented themselves for examination at a particular age period, and who had first contracted the disease at some previous age period. The totals of the columns and of the rows

give the distributions, according to the age at the time of examination, and to the age of onset of the disease respectively. Thus from Table I it is seen that three male patients were admitted between the ages of 35 and 39, in whom the first onset of the disease occurred between the ages of 25 and 29, and that the total number of males admitted between the ages of 35 and 39 was 12.

In interpreting these tables it must in the first place be remembered that the members of any particular age group including the asthmatics in that

Table I. *Males.*

Age at first attack	Age at time of examination															
	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	Totals
0–4	3	8	14	7	5	1	1	1	1	.	1	1	.	.	.	43
5–9	.	2	12	12	6	.	1	2	.	.	.	.	.	.	.	35
10–14	.	.	3	5	2	3	1	.	2	.	.	1	.	.	.	17
15–19	.	.	.	2	3	2	1	.	.	.	.	.	.	.	.	8
20–24	.	.	.	.	2	3	9	1	1	1	.	.	.	.	.	17
25–29	.	.	.	.	.	5	2	3	1	2	.	.	.	.	.	13
30–34	.	.	.	.	.	.	4	3	6	4	2	.	.	.	.	19
35–39	.	.	.	.	.	.	.	2	1	7	1	.	.	.	.	11
40–44	.	.	.	.	.	.	.	.	2	5	1	.	.	.	.	8
45–49	.	.	.	.	.	.	.	.	.	4	1	1	.	.	.	6
50–54	.	.	.	.	.	.	.	.	.	.	3	2	1	1	1	8
55–59	.	.	.	.	.	.	.	.	.	.	.	3	2	1	.	6
60–64	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	2
Totals	3	10	29	26	18	14	19	12	14	23	9	8	5	2	1	193

Table II. *Females.*

Age at first attack	Age at time of examination														
	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	Totals
0–4	2	3	8	6	5	3	.	2	.	.	.	.	.	.	29
5–9	.	5	5	4	2	4	.	.	1	.	.	.	1	.	22
10–14	.	.	.	2	2	3	.	1	.	1	.	.	.	.	9
15–19	.	.	.	2	8	5	5	3	.	.	1	.	.	.	24
20–24	.	.	.	.	3	7	4	1	4	2	.	.	1	.	22
25–29	.	.	.	.	.	5	7	1	3	.	1	.	.	1	18
30–34	.	.	.	.	.	.	3	5	4	1	.	1	.	.	14
35–39	.	.	.	.	.	.	.	2	3	1	1	1	1	.	9
40–44	.	.	.	.	.	.	.	.	1	3	.	1	.	.	5
45–49	.	.	.	.	.	.	.	.	.	2	1	2	.	.	5
50–54	.	.	.	.	.	.	.	.	.	.	3	4	1	1	9
55–59	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0
60–64	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1
Totals	2	8	13	14	20	27	19	15	16	10	7	9	5	2	167

group are constantly dying from diseases other than asthma, and secondly that the asthmatic individuals in that group are being reduced either by death from asthma or by the cure of their disease. We will in this analysis assume that the death rate of the asthmatic individuals from diseases other than asthma is the same as that of the population as a whole. If a difference does exist, for instance if the asthmatics are more liable to die from other diseases than the non-asthmatics—this extra death rate will appear in our analysis as an increase in the death rate due to asthma.

We shall denote as the *removal rate* this complex rate of removal of the asthmatics, which is arrived at after making allowance for the normal death

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

rate. It is to be realised that it includes not only recovery and death from asthma as mentioned above but also the effect of any increasing disinclination to report, such as might result from old age, or from habitude to the discomforts of the disease. A very important cause of the reduction may be the fact that the asthmatic individual learns how to avoid the factors which precipitate an attack and, either consciously or subconsciously, takes the necessary precautions.

We shall now make the tentative assumption that the removal rate is independent of the age at which the disease was contracted, and depends only upon the length of time during which the individual has suffered from the disease. We shall see later that, in fact, the figures do not appear to be inconsistent with this view.

We shall assume that the number of persons born r years ago was N_r per unit period, so that N_0 refers to the number born during the present period, which we shall assume is now closing. We shall later remove the limitation. The chance that an individual will contract asthma in the rth period from birth, assuming that he does not die of any other disease before that time, is P_r per unit period.

During any particular period then fresh cases are occurring amongst those within a particular period of age at the rate of $P_r N_r$ per unit period of time. Let us now concentrate our attention on the individuals who contracted asthma during a particular period of age, *e.g.* 35–39. At the present time these individuals may fall within the groups 35–39, 40–44, 45–49, or any higher age group. We will assume that the rate of recovery is α, so that, if R' is the number of people affected, $\frac{1}{R'}\frac{dR'}{d\theta} = -\alpha$ (where θ is the length of time the person has been affected). We will denote by $R_0, R_1, R_2, \ldots$, the number of individuals aged 35–39, 40–44, 45–49, etc., who would be found in the community as having acquired asthma between the ages of 35 and 39 if there were no deaths from other causes. It is clear that R_0 will be deficient as compared with R_1, R_2 and R_3, since the group to which it refers includes individuals who will ultimately acquire asthma during the period but have not yet become affected.

It is not difficult to express R_0, R_1 and R_2, ... in terms of NP and α, the removal rate. During the age period from 35 to 39, individuals are being affected with asthma at the rate NP per unit period, so that during a short space of time dT' the number of individuals between the ages τ and $\tau + d\tau$ who become affected is $NPd\tau dT'$. The examination is carried out at a time T, and we are considering individuals whose age then falls between t and $t + dt$. Clearly $T - T' = t - \tau = \theta$ the period during which the disease has continued. The number then of those who are infected between the times, T' and $T' + dT'$, and are of age between τ and $\tau + d\tau$, and who remain infected after a period θ, is

$$NPe^{-\int_0^\theta \alpha\, dx} d\tau dT'.$$

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

We shall now change the coordinates from τ and T' to τ and t, by means of the above relation, it being noted that T is constant, since it refers to the time at which observations are made. It follows that $d\tau dT' = d\tau dt$, and hence the number of age between t and $t + dt$, who were affected between τ and $\tau + d\tau$, is

$$NPe^{-\int_0^{t-\tau} \alpha dx} d\tau dt.$$

The total number who fall within the age group from t_0 to $t_0 + 1$, and who were infected when they were aged between τ_0 and $\tau_0 + 1$, is

$$\int_{\tau_0}^{\tau_0+1} \int_{t_0}^{t_0+1} NPe^{-\int_0^{t-\tau} \alpha dx} d\tau dt,$$

except in the case where $t_0 = \tau_0$, in which case the required number is

$$\int_{t_0}^{t_0+1} \int_{\tau_0}^{t} NPe^{-\int_0^{t-\tau} \alpha dx} dt d\tau,$$

since only individuals must be counted as affected if they are at or beyond the age at which infection occurs. It remains to calculate the values of these integrals for various values of t and τ. If we note that NP is a function of τ and varies between τ_0 and $\tau_0 + 1$, and apply the first theorem of the mean we find that

$$\int_{\tau_0}^{\tau_0+1} NPF(t, \tau)\, d\tau = (NP)_{\tau_0+\epsilon} \int_{\tau_0}^{\tau_0+1} F(t, \tau)\, d\tau,$$

where ϵ is less than unity.

If (NP) does not vary greatly during the course of a period, the mean value of (NP) during the period may be substituted for $(NP)_{\tau_0+\epsilon}$ without great error.

It may now be noted that the integral

$$\int_{\tau_0}^{\tau_0+1} \int_{t_0}^{t_0+1} e^{-\int_0^{t-\tau} \alpha dx} d\tau dt$$

depends on $t - \tau$, since α is a function of θ, that is of $t - \tau$ only. We shall denote by $\rho_1, \rho_2, \ldots$ the values of this integral when $t_0 - \tau_0 = 1, 2$, etc., and by ρ_0 the integral

$$\int_{t_0}^{t_0+1} \int_{\tau_0}^{t} e^{-\int_0^{t-\tau} \alpha dx} dt d\tau.$$

As the simplest assumption we put α equal to a constant. It will be shown below that this assumption agrees remarkably well with the data. It is easily found that

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

$$\rho_0 = \frac{1}{\alpha}\left[1 + \frac{e^{-\alpha} - 1}{\alpha}\right]$$

$$= \frac{1}{\alpha}\left[1 - \frac{e^{-\alpha/2}\sinh\frac{\alpha}{2}}{\frac{\alpha}{2}}\right];$$

$$\rho_1 = e^{-\alpha}\frac{\sinh^2\frac{\alpha}{2}}{\left(\frac{\alpha}{2}\right)^2};$$

$$\rho_2 = e^{-2\alpha}\frac{\sinh^2\frac{\alpha}{2}}{\left(\frac{\alpha}{2}\right)^2}; \text{ etc.}$$

It is now possible to draw up a scheme giving the numbers which theoretically ought to appear in the various compartments of Tables I and II. The final result is shown in Table III.

Table III.

Age at first attack	Age at time of examination: 0	1	2	3
0	$N_0P_0\rho_0$	$N_1\kappa_1P_0\rho_1$	$N_2\kappa_2P_0\rho_2$	$N_3\kappa_3P_0\rho_3$
1	—	$N_1\kappa_1P_1\rho_0$	$N_2\kappa_2P_1\rho_1$	$N_3\kappa_3P_1\rho_2$
2	—	—	$N_2\kappa_2P_2\rho_0$	$N_3\kappa_3P_2\rho_1$
3	—	—	—	$N_3\kappa_3P_3\rho_0$

In this table N_r refers to the number born r years ago, κ_r denotes the fraction of that number which is alive at the present time. We shall denote by κ_r^{τ} the fraction of this particular group who were living after τ periods. Hence N of our previous paragraph $= N_r\kappa_r^{\tau}$. As explained in the above paragraph, for NP we should take the value corresponding to some particular age occurring within the period τ to $\tau + 1$, but in most cases it will be sufficiently accurate to take $N_r\kappa_r^{\tau}P_{\tau}$, where P_{τ} is the mean value of P for the age period. This is the number of asthmatics who become ill during this age period, whose age now lies between r and $r + 1$. These will have been reduced in number in two ways, first by the action of diseases other than asthma, which is supposed to act on asthmatics as on other individuals, who are reduced from $N_r\kappa_r^{\tau}$ to $N_r\kappa_r$, and secondly by the removal of asthmatics, which results in the introduction of the factors ρ_0, ρ_1, ρ_2, etc. Hence the figures which ought to appear in the compartments are

$$N_r\kappa_r^{\tau}P_{\tau}.\frac{\kappa_r}{\kappa_r^{\tau}}.\rho_{r-\tau} = N_r\kappa_rP_{\tau}\rho_{t-\tau}.$$

Also $\frac{N_r\kappa_r}{\Sigma N_r\kappa_r}$ represents the fraction of the present population who fall within the age period r, and hence the relative values of $N_r\kappa_r$ are readily found from the census figures.

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

These have been kindly supplied to us by the Registrar-General for Scotland, and are given in Table IV. They are calculated for the year 1926, which is approximately the year about which our observations are centred. The various P's and ρ's are unknown, and the problem is to calculate them

Table IV. *Estimated population of Scotland June 30th*, 1926*, *to nearest thousand.*

	Males	Females	Both sexes
Under 1 year	49	48	97
1–4	198	193	391
5–9	221	217	439
10–14	230	227	458
15–24	445	453	898
25–34	337	396	733
35–44	280	332	612
45–54	267	289	556
55–64	192	207	399
65–74	103	124	228
75–84	29	47	76
85 and over	3	8	11
All ages	2354	2542	4897

* Based on 1921 Census populations, births, deaths at each age, and estimated net emigration at each age.

from Tables I and II. It is at once clear that if we divide each column of Tables I and II by the appropriate value of $\frac{100N_r\kappa_r}{\Sigma N_r\kappa_r}$ derived from the census tables, we shall obtain a series of values which would correspond to the theoretical scheme given in Table V, where λ is a constant whose value is

Table V.

		Age at time of examination			
		0	1	2	3
Age at first attack	0	$\lambda P_0\rho_0$	$\lambda P_0\rho_1$	$\lambda P_0\rho_2$	$\lambda P_0\rho_3$
	1	—	$\lambda P_1\rho_0$	$\lambda P_1\rho_1$	$\lambda P_1\rho_2$
	2	—	—	$\lambda P_2\rho_0$	$\lambda P_2\rho_1$
	3	—	—	—	$\lambda P_3\rho_0$

$N/100$ and N is the number of the susceptible population from which the patients were drawn, and which cannot be determined. If then we add the rows of this table we find values of

$$\lambda P_0\,(\rho_0 + \rho_1 + \rho_2 + \ldots),$$
$$\lambda P_1\,(\rho_0 + \rho_1 + \rho_2 + \ldots),$$
$$\lambda P_2\,(\rho_0 + \rho_1 + \rho_2 + \ldots), \text{ etc.}$$

which are proportional to P_0, P_1, P_2, etc.

Since λ is unknown, only the relative values of the P's are of significance. If, however, the P's are taken such that $P_0 + P_1 + P_2 + \ldots = 1$, this is equivalent to considering all those who would ultimately suffer from asthma (assuming that no deaths occurred from other diseases) as a separate group, the P's representing the fraction of this group which would contract asthma at various age periods.

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

No information however is available as to the proportion which this group bears to the total population, and this can only be calculated on the basis of other assumptions more or less plausible.

If we sum the diagonals of Table V we obtain the values

$$\lambda\rho_0 (P_0 + P_1 + P_2 + \ldots),$$
$$\lambda\rho_1 (P_0 + P_1 + P_2 + \ldots),$$
$$\lambda\rho_2 (P_0 + P_1 + P_2 + \ldots), \text{ etc.},$$

so that the ratios

$$\frac{\rho_1}{\rho_0}, \frac{\rho_2}{\rho_0}, \frac{\rho_3}{\rho_0} \ldots$$

are also determined.

When then the above theory is applied to Tables I and II, the steps are first to divide each number in the table by the appropriate $N_r\kappa_r$ derived from the census figures. Since only the ratios are important it is convenient to take the percentage which the rth age group bears to the population as derived from Table IV. The rows and diagonals of this new table are then summed. The sums of the rows are then divided by their total, whilst the sums of the diagonals are divided by the sum of the first diagonal. The values of ρ_s/ρ_0, and of the P's thus found, are shown in Tables VI and VII.

Table VI.

Stage of disease in periods of 5 years	ρ_s/ρ_0 values. Females		Males	
	Extracted	Calculated	Extracted	Calculated
0	1·00	1·00	1·00	1·00
1	1·66	1·52	1·14	1·44
2	1·10	1·06	1·48	0·93
3	0·88	0·74	0·62	0·60
4	0·60	0·52	0·42	0·38
5	0·38	0·36	0·08	0·26
6	0	0·26	0·14	0·16
7	0·20	0·18	0·02	0·10
8	0·16	0·12	0·04	0·07
9	0	0·08	0·04	0·04
10	0	0·06	0·04	0·03
11	0·06	0·04	0·04	0·02
	6·04	—	5·06	—
		$\alpha = 0{\cdot}360$		$\alpha = 0{\cdot}437$

Table VII.

Ages	P values		p values	
	Females	Males	Females	Males
0–4	0·141	0·173	0·141	0·174
5–9	0·115	0·136	0·134	0·165
10–14	0·047	0·075	0·063	0·109
15–19	0·127	0·032	0·182	0·052
20–24	0·129	0·081	0·226	0·139
25–29	0·112	0·066	0·254	0·131
30–34	0·088	0·114	0·267	0·261
35–39	0·066	0·070	0·274	0·217
40–44	0·037	0·052	0·206	0·206
45–49	0·039	0·041	0·281	0·205
50–54	0·085	0·083	0·850	0·522
55–59	0	0·056	0	0·737
60 and over	0·015	0·020	1·000	1·000

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

It has been shown above that on the assumption that α is constant

$$\frac{\rho_s}{\rho_0} = \frac{e^{-s\alpha}\dfrac{\sinh^2\dfrac{\alpha}{2}}{\left(\dfrac{\alpha}{2}\right)^2}}{\dfrac{1}{\alpha}\left[1 - e^{-\alpha/2}\dfrac{\sinh\dfrac{\alpha}{2}}{\dfrac{\alpha}{2}}\right]}.$$

We have then to find the value of α which will best fit any particular set of ρ's found from statistics. It is not difficult to show that the approximate value of α is in the neighbourhood of 0·3, and so $\dfrac{\sinh \alpha/2}{\alpha/2}$ differs but slightly from unity.

Thus

$$\frac{\rho_s}{\rho_0} = \frac{e^{-s\alpha}}{\frac{1}{\alpha}(1 - e^{-\alpha/2})}$$

$$= \frac{2e^{-s\alpha}}{e^{-\alpha/4}\dfrac{\sinh\dfrac{\alpha}{4}}{\dfrac{\alpha}{4}}}$$

$$= 2e^{-(s-\frac{1}{4})\alpha}.$$

Also

$$\left(1 + \frac{\rho_1}{\rho_0} + \frac{\rho_2}{\rho_0} + \ldots\right) = 1 + 2e^{-\frac{1}{4}\alpha} + 2e^{-1\frac{1}{4}\alpha} + 2e^{-2\frac{1}{4}\alpha} + \ldots$$

$$= 1 + \frac{2e^{-\frac{1}{4}\alpha}}{1 - e^{-\alpha}}$$

$$= 1 + \frac{e^{-\alpha/4}}{\sinh\dfrac{\alpha}{2}}.$$

The value of α can thus be readily obtained, if necessary by interpolation. In this way it is found that for males $\alpha = 0{\cdot}437$ and for females $\alpha = 0{\cdot}360$. The corresponding values of ρ_s/ρ_0 are given in Table VI and in Figs. 1 and 2, and it will be seen that there is good agreement with the series found from the statistics, particularly in the case of the females. Since the above figures are calculated on the basis of a five-year period, the corresponding values taking one year as unit would be 0·087 and 0·072 respectively. In other words there is a removal rate amongst the asthmatics of about 8 per cent. per year. As explained above several causes may contribute to this removal rate, and it is not possible from the statistics to decide as to their relative importance.

The values of P for males and females may now be considered (Table VII). It will be seen that in both cases there is a distinct fall to a minimum then a rise to a maximum followed by a fall, a definite check in the fall occurring

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

at age 50. It is to be remembered that the absolute P values for males and females are not directly comparable, but since the total numbers for males and females are not very dissimilar there is probably no great error involved in taking them as based on a common unit. The minimum and first maximum occur at an earlier period in the case of the females than in the case of the

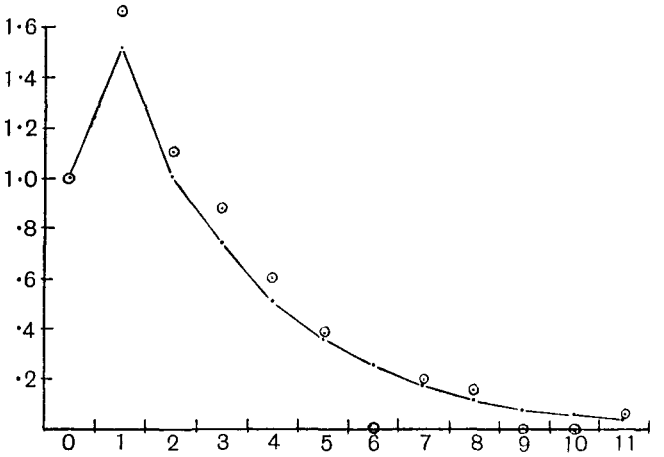

Fig. 1. Chart of ρ_s/ρ_0 values for females. The rings represent values extracted from Table II as explained in the text. The dots, connected for clearness by straight lines, are the values calculated on the assumption that the rate of removal a has a constant value of 0·360.

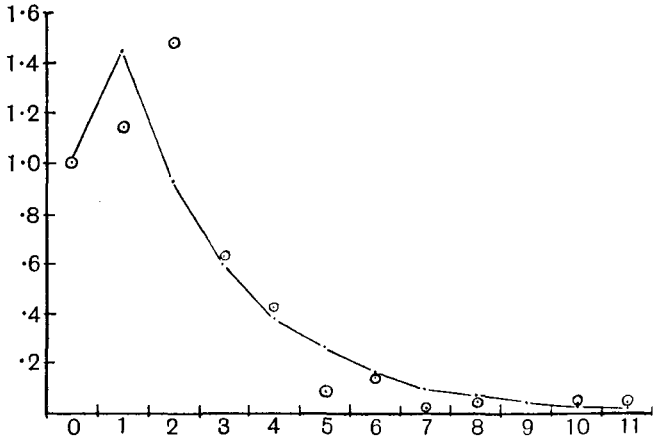

Fig. 2. Chart of ρ_s/ρ_0 values for males. The rings represent values extracted from Table I. The dots, connected by straight lines, are the values calculated on the assumption that the rate of removal a has a constant value of 0·437.

males. In other respects the curves are not markedly dissimilar. It is to be remembered that the later values (over 50) are based on very small figures and therefore their error is larger. The maximum at the age 50 in both curves may be due to a general tendency to state dates and ages in round numbers.

The original tables may now be recalculated from the P and ρ values which we have arrived at. The recalculated female figures are shown in

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

Table VIII. *Females.*

Age at time of examination

Age at first attack	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85 and over	Totals
0–4	5·31	7·25	5·29	3·66	2·51	1·62	1·08	0·69	0·43	0·27	0·18	0·10	.	.	.	.	.	.	28·39
5–9	.	3·89	6·17	4·27	2·91	1·91	1·22	0·82	0·52	0·32	0·20	0·12	0·06	.	.	.	.	.	22·41
10–14	.	.	1·68	2·53	1·71	1·11	0·73	0·45	0·31	0·20	0·12	0·06	0·03	0·03	.	.	.	.	8·96
15–19	.	.	.	4·47	6·64	4·31	2·79	1·80	1·17	0·79	0·51	0·27	0·15	0·08	0·03	.	.	.	23·01
20–24	.	.	.	.	4·46	6·30	4·08	2·62	1·73	1·13	0·76	0·41	0·22	0·10	0·06	0·02	.	.	21·89
25–29	.	.	.	.	.	3·61	5·08	3·26	2·14	1·41	0·91	0·52	0·30	0·15	0·07	0·03	0·01	.	17·49
30–34	.	.	.	.	.	.	2·64	3·68	2·42	1·59	1·04	0·57	0·33	0·17	0·08	0·03	0·01	.	12·56
35–39	.	.	.	.	.	.	.	1·80	2·58	1·68	1·10	0·61	0·34	0·19	0·09	0·04	0·01	.	8·44
40–44	.	.	.	.	.	.	.	.	0·93	1·32	0·86	0·47	0·27	0·13	0·07	0·03	0·01	.	4·09
45–49	.	.	.	.	.	.	.	.	.	0·96	1·36	0·75	0·43	0·22	0·11	0·05	0·01	0·01	3·90
50–54	.	.	.	.	.	.	.	.	.	.	1·92	2·31	1·33	0·68	0·34	0·16	0·05	0·03	6·82
55–59	.	.	.	.	.	.	.	.	.	.	.	0	.	.	.	.	.	.	0
60–64	.	.	.	.	.	.	.	.	.	.	.	.	0·16	0·18	0·09	0·04	0·01	0·01	0·49
Totals	5·31	11·14	13·14	14·93	18·23	18·86	17·62	15·12	12·23	9·67	8·96	6·19	3·62	1·93	0·94	0·40	0·11	0·05	158·45

Table IX. *Males.*

Age at time of examination

Age at first attack	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85 and over	Totals
0–4	10·05	12·95	8·69	5·49	3·39	2·21	1·17	0·25	0·19	0·11	0·07	0·04	.	.	.	.	.	.	44·61
5–9	.	7·03	10·51	6·64	4·18	2·52	1·48	0·73	0·20	0·14	0·09	0·04	0·03	.	.	.	.	.	33·59
10–14	.	.	4·03	5·67	3·56	2·21	1·19	0·65	0·38	0·11	0·07	0·04	0·02	0·01	.	.	.	.	17·94
15–19	.	.	.	1·67	2·33	1·45	0·79	0·40	0·20	0·16	0·05	0·03	0·02	0·01	0	.	.	.	7·17
20–24	.	.	.	.	4·16	5·71	3·17	1·64	0·97	0·63	0·37	0·09	0·05	0·03	0·01	0	.	.	16·83
25–29	.	.	.	.	.	3·26	4·01	2·09	1·26	0·77	0·49	0·26	0·07	0·04	0·01	0	0	.	12·26
30–34	.	.	.	.	.	.	4·81	5·57	3·35	2·08	1·23	0·72	0·37	0·09	0·03	0·01	0	0	18·26
35–39	.	.	.	.	.	.	.	2·39	3·21	1·99	1·21	0·66	0·37	0·18	0·03	0·01	0	0	10·05
40–44	.	.	.	.	.	.	.	.	1·65	2·29	1·38	0·77	0·39	0·22	0·07	0·01	0	0	6·78
45–49	.	.	.	.	.	.	.	.	.	1·25	1·68	0·93	0·49	0·25	0·09	0·03	0	0	4·72
50–54	.	.	.	.	.	.	.	.	.	.	2·37	2·93	1·53	0·81	0·25	0·10	0·03	0·01	8·03
55–59	.	.	.	.	.	.	.	.	.	.	.	1·39	1·62	0·86	0·27	0·10	0·03	0·02	4·29
60–64	.	.	.	.	.	.	.	.	.	.	.	.	0·40	0·48	0·15	0·06	0·02	0·01	1·12
Totals	10·05	19·98	23·23	19·47	17·62	17·36	16·62	13·72	11·47	9·53	9·01	7·90	5·36	2·98	0·91	0·32	0·08	0·04	185·65

Table VIII and the totals of the rows and columns correspond to the original figures (Table II) referring to age at onset and age at time of examination. The calculated and original curves follow each other closely both for figures referring to age at time of examination (Fig. 3) and for those referring to age at onset. This is not surprising with regard to the latter series since they depend very directly upon the P values.

An entirely satisfactory method for testing the goodness of fit of a table, such as Table II, does not seem to be at present available. The simplest process which suggests itself is to apply the χ^2 test to the table as a whole by summing the contributions from the separate compartments, each referring to a particular age at time of examination, and to a particular age of onset.

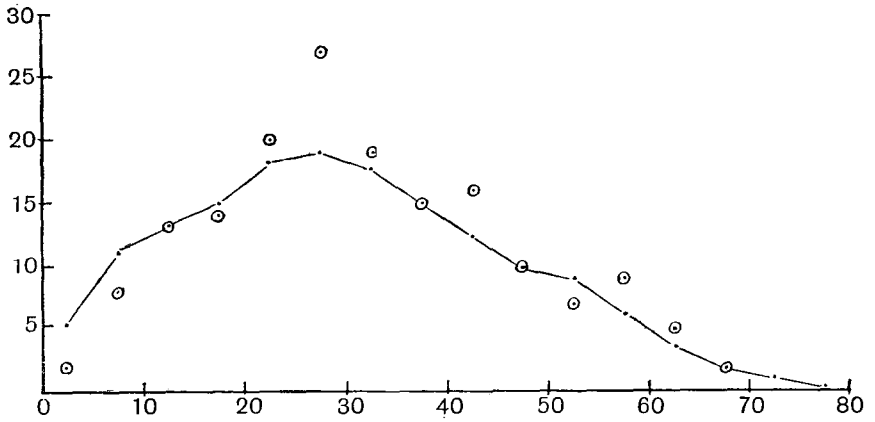

Fig. 3. Numbers of females who presented themselves for examination at various ages. The rings are the observed values tabulated in Table II. The dots, connected for clearness by straight lines, are the values obtained on recalculation from the P values tabulated in Table III, and from the values of ρ_s/ρ_0 tabulated in the second column of Table VI, *i.e.* on the assumption that the rate of removal a has a constant value of 0·360.

The difficulty arises however that many of the observed frequencies are quite small. In the case of those cells where the theoretical frequency is less than unity it is clearly unsatisfactory to make use of the corresponding figures without modification, and the simplest device is that usually adopted, namely, to add up groups of such figures, and combine them in a single compartment. None of the observed frequencies however exceeds 8, and so there appears to be little justification for the assumption made in the development of the χ^2 test, that the distribution of the observed frequency in any cell about the theoretical for that particular cell, is a normal one. If there are n cells and the theoretical value for each cell is high, the distribution of χ^2, assuming there are no restrictions imposed, is given by $\kappa\chi^{n-1}e^{-\chi^2/2}d\chi$, where κ is adjusted so that the integral of the above expression from $\chi = 0$ to $\chi = \infty$ is equal to unity. The distribution of χ^2 is therefore independent of the theoretical values for the separate cells, and is solely dependent upon n. If, however, the theoretical values are quite small, and n is also small, it is not difficult to show that

the distribution of χ^2 is markedly discontinuous, and may differ considerably from the above value. The tendency appears to be for the occurrence of an unexpectedly large number of the higher values of χ^2, and a corresponding deficiency of the smaller values. If now the number of cells, instead of being small, is very considerable, but the theoretical values for each cell are still quite low, the distribution of χ^2 will become more nearly continuous in spite of the marked discontinuity of the contributions from the separate cells, but it is difficult to say without further investigation what particular form it assumes. In our present state of knowledge it seems justifiable as a rough method to apply the χ^2 test to such a table after grouping the frequencies for which the theoretical expectations are less than unity.

The value of χ^2 obtained is a measure of the discrepancy between the theoretical and observed frequencies, but will naturally depend upon the methods used in calculating the theoretical values, as well as upon the laws which are assumed. If the theoretically best methods are used in the calculation of these theoretical values, the value of χ^2 will measure the agreement of the data with the laws which have been assumed, and this will be approximately true when χ^2 is a minimum, or when χ^2 satisfies the condition of maximum likelihood (R. A. Fisher, 1928). It is to be noted that in consequence of the above considerations a difficulty presents itself whenever any grouping of the frequencies has to be carried out before applying the test of goodness of fit. This is because the frequencies so grouped will in all probability not even approximately satisfy the condition that χ^2 is a minimum, or any other satisfactory criterion. This is particularly evident if the grouping is carried out to such an extent that the number of cells is reduced to the number of constants. Clearly the best fit would be attained when the constants were so adjusted that each of these cells would give absolute agreement with the observed, and this agrees with the circumstance that when, as in this case, $n' = 1$ (since the number of constants is equal to the number of cells), any deviation from absolute agreement is quite impossible.

These considerations make clear the reason for the difficulty which is encountered when an attempt is made to apply the χ^2 test to the sums of the rows, or the sums of the columns. If it were possible to do so this would constitute a very convenient test of the hypothesis, since the frequencies at most of the ages are fairly large, and the distributions refer to conceptions of a simple and fundamental nature. In the case of the sums of the rows we do in fact find that the agreement is almost perfect; and since the number of constants is equal to the number of rows (for this purpose α is not to be taken as a constant, since changes in α are practically without effect upon the totals in any row) this merely means that the calculations have been carried out in a way quite satisfactory as far as concerns the totals of the rows.

If we now consider the sums of the columns it will be clear that the P's are not related to these totals in the same direct way as they are to the sums of the rows, but that the constant α now plays an important part in

determining the distribution. However, the relation between the constants and the actual frequencies is a somewhat complicated one, and when these totals alone are considered, the constants are by no means such as to give the best agreement between the observed and the calculated figures. It therefore does not seem possible to apply the χ^2 test to these totals in a way likely to yield information as to the goodness of fit of the calculated frequencies.

The method of testing open to the least grave objections appears to be the application of χ^2 to the table as a whole, after those cells have been grouped in which the theoretical value is very small. It is always to be borne in mind, however, that this method can at best be expected to yield only approximate results, and that the criterion for satisfactory goodness of fit should not be as strict as that ordinarily employed, for the reasons outlined above.

Applying this calculation to the above figures for females we find $\chi^2 = 45{\cdot}6$, $n = 44$, $n' = 45$. (There were 57 cells and the number of constants involved was 13, hence $n = 44$.) From Elderton's table it is found that $P = 0{\cdot}33$. Thus there are approximately 33 chances in 100 that a discrepancy equal to, or greater than, the above would occur as the result of random sampling. It therefore appears, so far as this method can show, that the assumptions made are consistent with the data.

The following test appears to be more searching, and leads to a similar result. If we have m groups each containing n observations, and the theoretical value for the members of the same group is s, then if we calculate χ^2 for the n observations of the same group we shall have m values of χ^2, and these ought to be distributed with a frequency $N\chi^{n-1}e^{-\chi^2/2}d\chi$. If the theoretical value s is actually deduced from the n observations of the group, then the

χ^2	Expected	Observed
∞		
	0·57	0
6·635		
	0·57	0
5·412		
	1·71	0
3·841		
	2·85	2
2·706		
	5·70	9
1·642		
	5·70	7
1·074		
	11·40	12
0·455		
	11·40	11
0·148		
	5·70	7
0·0642		
	11·40	9
0		

distribution will be $N\chi^{n-2}e^{-\chi^2/2}d\chi$. In our problem $n = 1$. The theoretical value is deduced not from single observations, but from the observations as a whole, using 13 constants. One would therefore expect the distribution of χ^2 to be given by $Ne^{-\chi^2/2}$, although some deviation from this law would not be sur-

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

prising, in view of the fact that the theoretical values are not absolutely independent of the data in so far as they are functions of 13 constants determined from the 57 observed frequencies. As however the total number of observations is comparatively small, namely 57, considerable deviations will occur as the result of random sampling, and therefore it is probably sufficiently approximate to take the above law as describing the distribution of χ^2 in a satisfactory manner. To apply this method we calculate $\frac{(x-s)^2}{s}$ for each cell. The table on p. 342 gives the frequency distribution of the 57 observations as actually obtained, and also the theoretical distribution of these 57 observations according to the law.

The agreement of the theoretical and the observed frequencies in the various ranges of χ^2 may itself be tested by means of the χ^2 method. In this case the only constant entering into the distribution of χ^2 is the total number, and thus we calculate χ^2 in the usual way, and find $P = 0{\cdot}65$ from Elderton's tables, taking $n' = 9$. The actual distribution is therefore in excellent agreement with the calculated.

When we consider the table for males (Table I), it is at once seen that the distribution of the figures is by no means as regular as in the case of the females, and therefore it is unlikely that any law will give equally satisfactory results. We shall however treat the table by exactly similar methods, and shall then consider in what particular respects, if any, the discrepancies appear. The calculated figures are shown in Table IX (and Fig. 4), and the

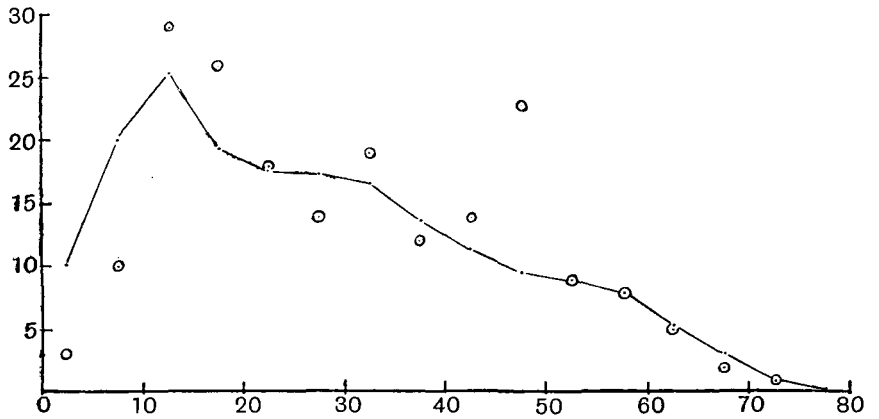

Fig. 4. Numbers of males who presented themselves for examination at various ages. The rings are the observed values tabulated in Table I. The dots, connected by straight lines, are the values obtained on recalculation from the P values tabulated in Table VII, and from the values of ρ_s/ρ_0 tabulated in the fourth column of Table VI, *i.e.* on the assumption that the rate of removal α has a constant value of 0·437.

total value of χ^2 is 84·4 after grouping the very small frequencies as was explained above. In this case the number of constants employed was 14, $\therefore n = 56 - 14 = 42$, and $n' = 43$. Thus $P = 1{\cdot}49 \times 10^{-4}$.

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

It is thus not likely that this result will be obtained as the result of errors of sampling. We may now split up χ^2 as before. The result is shown in the following table:

χ^2	Expected	Observed
∞	0·56	2
6·635	0·56	2
5·412	1·68	2
3·841	2·80	3
2·706	5·60	5
1·642	5·60	6
1·074	11·20	10
0·455	11·20	14
0·148	5·60	9
0·642	11·20	3
0		

When we apply the χ^2 test to this distribution we find $P = 0{\cdot}035$. This value of P, although quite consistent with chance, suggests that there may possibly be some discrepancy between calculation and theory. The value $P = 1{\cdot}49 \times 10^{-4}$ deduced from the total value of χ^2 indicates a very low probability. The apparent anomaly seems to be due to the fact that amongst the four values of χ^2 greater than 5·412, two at least are extremely high, and very improbable. It is clear that if these values had been somewhat smaller they would affect the value of P deduced from the total χ^2, but not that deduced from the above table. There is also a marked deficiency of very small values of χ^2, suggesting that apart from the few extreme aberrations, there is also a general slight lack of conformity between the observed and the theoretical figures; but too much emphasis must not be laid upon this since, as explained above, the fact that the theoretical values of the frequencies are quite small would probably tend to cause a deficiency of very small contributions to χ^2. It may again be emphasised that the P values have been obtained by methods which can be justified only as approximations, and that therefore they are subject to errors which may be quite considerable. The general conclusion is that some discrepancy appears to exist between the calculated and the observed values, which is not likely to be the result of random sampling, but which is mainly confined to two or three particular groups, and that the laws suggested represent in a general way, at least, the actual course of events.

As explained above the P values (Table VII) give the fraction of susceptible persons in a given number who contract the disease for the first time during a particular age period. If now p is the probability that a person of a particular age, who ultimately will develop the disease, contracts the disease during the age period in which he finds himself, then

https://doi.org/10.1017/S0022172400010056 Published online by Cambridge University Press

$$P_r = p_r (1 - p_0)(1 - p_1) \ldots (1 - p_{r-1}),$$

or

$$p_r = \frac{P_r}{1 - (P_0 + P_1 + \ldots P_{r-1})}.$$

This latter formula is self evident, if it be noted that P_r gives the number who contract the disease during the rth period divided by the number of susceptibles originally present in the group, whereas p_r is equal to the same number divided by the number of persons who have not yet contracted the disease, and that if N is the original number, the number who have not yet contracted the disease is $N - NP_0 - NP_1 - \ldots$. The values of p_r as calculated by the above formula are given in Table VII. As before it must be remembered that the error for age periods above 50 is very considerable.

In both male and female curves (Fig. 5) it will be seen that there is a dip followed by a rise, and then a portion which is almost flat. Both curves would

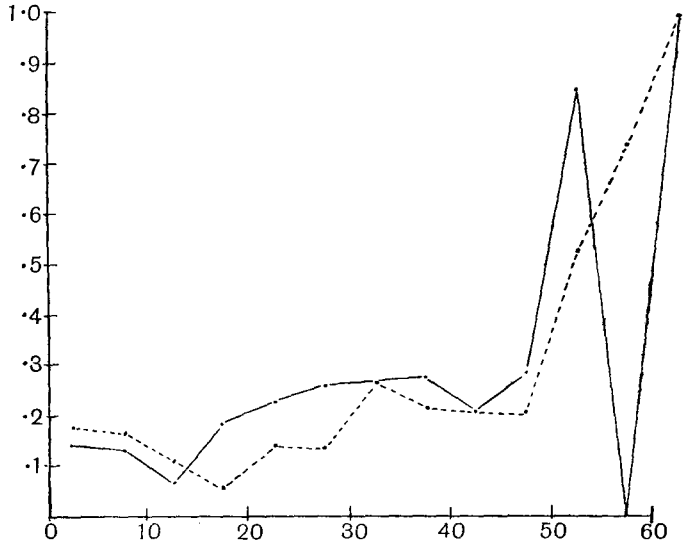

Fig. 5. Chart of the p values for different ages, extracted as explained in the text. The points connected by continuous lines refer to females, and those connected by interrupted lines to males.

seem to rise in the later periods but the error here is particularly great, owing to the smallness of the figures, and also to the difficulties introduced by the fact that the P values are regarded as changing discontinuously from period to period. The chief difference between the curves for the two sexes is that the male curve lags behind the female curve in its various characteristics.

The p_r values tell us that of every 1000 people aged for example 45, and of the female sex, who are susceptible to asthma, but have not yet been affected, 281 will acquire the disease during the next five years. It will be seen that apart from the drop during adolescence, this probability gradually increases with increasing age.

In conclusion we desire to express our thanks to Prof. Major Greenwood for his helpful criticism.

SUMMARY.

A statistical analysis has been made of the figures collected during the last eight years in the Edinburgh area, relating to asthmatic patients who presented themselves for examination in certain departments of the Edinburgh Royal Infirmary. The figures for males and females are considered separately, and refer to the numbers of individuals who presented themselves at various ages, and in whom the disease first manifested itself at various earlier ages. In all, complete records were obtained of 193 males and 167 females. For convenience the ages are grouped in five-year periods.

It has been shown that the figures in the case of females agree well with the assumption that asthmatics, no matter at what age they are first affected, gradually cease to suffer from the disease, or at least fail to report for examination, at a rate which is constant and independent of age. In the case of males the agreement between theory and observation is not as good as in the case of the females, but much of the discrepancy results from two or three obviously aberrant figures. Of every 100 asthmatics at any age it would appear that approximately 7 females, or 8 males as the case may be, are transferred to a non-complaining category every year. The absence of complaint may be due to death from asthma, or recovery, or to some other cause, as for example, that by repeated experience the individual learns how to avoid conditions which precipitate an attack.

On this assumption the relative rates of incidence (P) of the disease at various ages have been calculated for males and for females. Both curves fall during adolescence, rise to a maximum, and then gradually fall, but the male curve lags behind the female curve. From these figures the rates of attack amongst the potential asthmatics (p) have been calculated both for males and for females. The curves fall during adolescence, then rise, remaining practically constant until about 45 years of age, when a rise occurs towards unity. For the further analysis of these curves larger numbers of figures will be necessary, in order that confidence may be placed in the results.

(*MS. received for publication* 25. VII. 1929.—Ed.)