A not-well-known Lionel Penrose - Ronald Fisher collaboration

James Hanley
Department of Epidemiology, Biostatistics and Occupational Health
McGill University, Montréal, Québec, Canada

Webinar
September 24, 2025

Sponsored by CHANCE magazine and the ASA History of Statistics Interest Group

OUTLINE

- Down Syndrome: what it is; epidemiology

- 1933 (Mother's age? Father's Age?) CHANCE 2025
  - Penrose J Gen.,1933: Relative Effects of Parental Ages - data, two analyses, no mention of input from Fisher
  - Sudoku 2014-2023: Reconstructing $42 \times 31 \times 2$ Table
    – Initial surprises
    – What does/does not vary across multiple solutions
    – Sufficiency, various fits

- 1934 (Mother's age? Birth Order?) Biometrika 2024
  - Penrose Proc. R. Soc. Lond. B, 1934 results
  - Penrose Ann. Eugen,1934 Fisher / cond'nl logistic regrn.
  - Archives UCL/Adelaide Penrose-Fisher collaboration

Material related to Biometrika 2024 article on the conditional logistic regression fitted by Lionel Penrose and Ronald Fisher in 1934.

1934 Articles

* The relative aetiological importance of birth order and maternal age in mongolism.
  Penrose LS. Proceedings of the Royal Society B 115 431-450.
* A method of separating the relative aetological effects of birth order and maternal age, with specific reference to mongolian imbecility.
  Penrose LS. Annals of Eugenics 6 108-132.

Data set

* Penrose's 1934 Data: 217 sibships with at least one child with DS (.csv)
  The dataset was reconstructed from the appendices to the two 1934 articles.
  It has information on 224 affected children (the no. reported in the articles),
  but on 80**6** unaffected children (rather than the 80**7** reported)

Archives

* Birth Order and Down's syndrome Correspondence at Wellcome Library
* Penrose Papers at UCL Digital Collections
* R A Fisher Correspondence Files at University of Adelaide
* Penrose's diary account of the family tour of North America in 1958
  before the 10th International Congress of Genetics at McGill University, and the conferring of an honourary doctorate (last few slides)
* Penrose's diary account of his 1964 visit to the USA, where (in NYC) he received the Joseph P. Kennedy Jr. Foundation's International Award from LBJ
* First Awards ceremony (1962), where the awards were presented by JFK.

Material related to

"The first conditional logistic regression and the first nested case-contol study"

McGill Biostatistics Seminar Series, 2024.11.27

Presentation

video | 'slides' lyrics (these include some material not in video/audio)

Background:

* Reviews: Cole-1979 Liddell-1988 Breslow-1996 Doll-2001 Breslow-2007
* Articles:

Webinar presentation

"A not-well-known Lionel Penrose - Ronald Fisher collaboration"

Sponsored by CHANCE magazine and the ASA History of Statistics Interest Group

September 24, 2025

video, to be added | 'slides' | lyrics

Why we are telling these two stories ...

- Epidemiology/genetics/statistics/optimization/history
- Opportunity to reflect on
  - 90 years of growth in statistical methods/computing
  - what we have forgotten along the way
  - sufficient statistics and data privacy/disclosure
  - ‘let the analysis fit the design’
  - ML fitting of a conditional logistic regression in 1934

Down syndrome

87 languages

Article Talk Read View source View history Tools

From Wikipedia, the free encyclopedia

Down syndrome or **Down's syndrome**, also known as **trisomy 21**, is a genetic disorder caused by the presence of all or part of a third copy of chromosome 21.[3] It is usually associated with developmental delays, mild to moderate intellectual disability, and characteristic physical features.[1][12] There are three types of Down syndrome, all with the same features: Trisomy 21, the most common type; Mosaic Down syndrome, and Translocation Down syndrome.[13][14]

The parents of the affected individual are usually genetically normal.[15] The probability increases from less than 0.1% in 20-year-old mothers to 3% in those of age 45.[4] The extra chromosome is provided at conception as the egg and sperm combine.[16] A very small percentage of 1-2% gets the additional chromosome in the embryo stage and it only impacts some of the cells in the body; this is known as Mosaic Down syndrome.[17][18] Usually, babies get 23 chromosomes from each parent for a total of 46, whereas in Down syndrome, a third 21st chromosome is attached.[18] It is believed to occur by chance, with no known behavioral activity or environmental factor that changes the probability.[2] Down syndrome can be identified during pregnancy by prenatal screening, followed by diagnostic testing, or after birth by direct observation and genetic testing.[6] Since the introduction of screening, Down syndrome pregnancies are often aborted (rates varying from 50 to 85% depending on maternal age, gestational age, and maternal race/ethnicity).[19][20][21]

There is no cure for Down syndrome.[22] Education and proper care have been shown to provide good quality of life.[7] Some children with Down syndrome are educated in typical school classes, while others require more specialized education.[8] Some individuals with Down syndrome graduate from high school, and a few attend post-secondary education.[23] In adulthood, about 20% in the United States do paid work in some capacity,[24] with many requiring a sheltered work environment.[8] Support in financial and legal matters is often needed.[10] Life expectancy is around 50 to 60 years in the developed world, with proper health care.[9][10] Regular screening for health issues common in Down syndrome is recommended throughout the person's life.[9]

Down syndrome

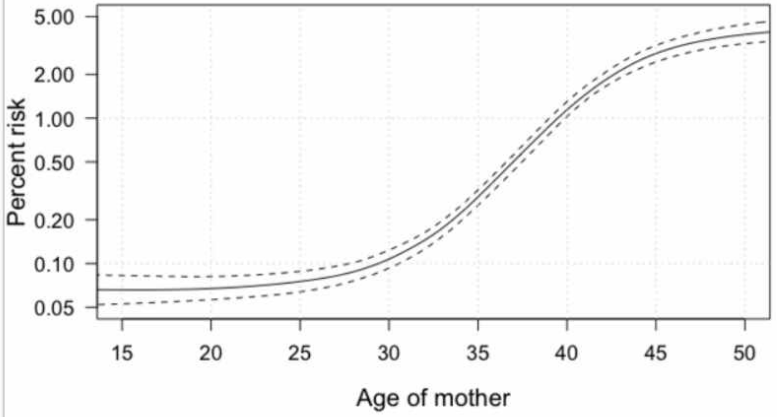
Other names Down's syndrome, Down's, trisomy 21

An eight-year-old boy displaying characteristic facial features of Down syndrome

Specialty Medical genetics, pediatrics

Epidemiology

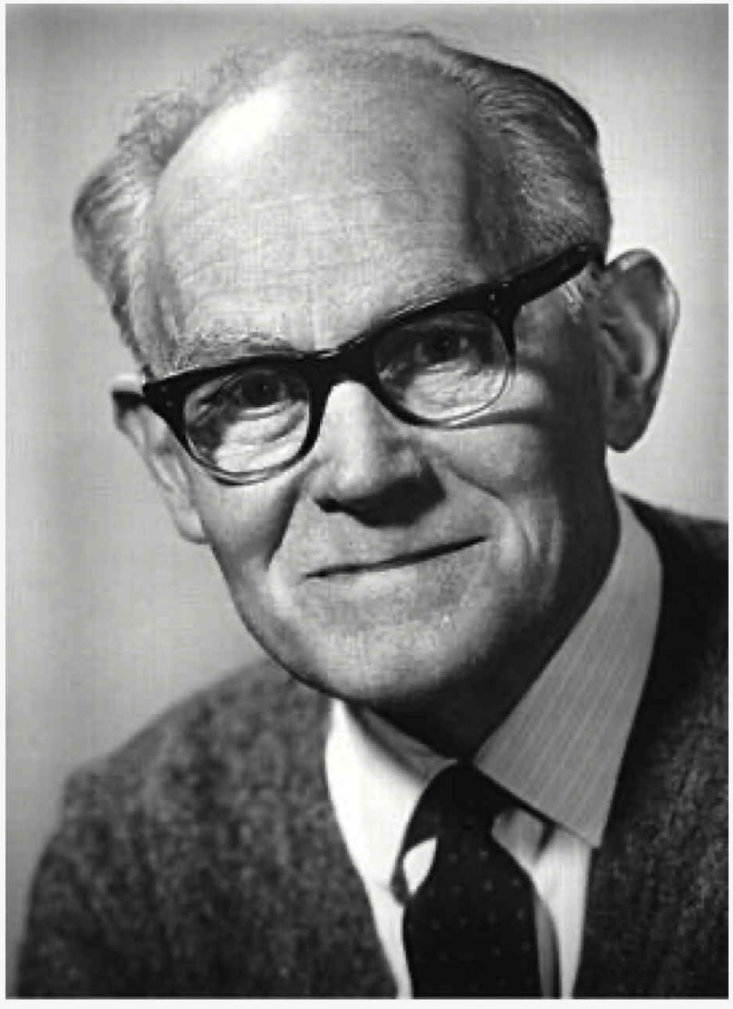
Down syndrome is the most common chromosomal abnormality in humans.[9] Globally, as of 2010, Down syndrome occurs in about 1 per 1,000 births[1] and results in about 17,000 deaths.[132] More children are born with Down syndrome in countries where abortion is not allowed and in countries where pregnancy more commonly occurs at a later age.[1] About 1.4 per 1,000 live births in the United States[133] and 1.1 per 1,000 live births in Norway are affected.[9] In the 1950s, in the United States, it occurred in 2 per 1,000 live births with the decrease since then due to prenatal screening and abortions.[92] The number of pregnancies with Down syndrome is more than two times greater with many spontaneously aborting.[10] It is the cause of 8% of all congenital disorders.[1]



The risk of having a Down syndrome pregnancy in relation to a mother's age[4]

Maternal age affects the chances of having a pregnancy with Down syndrome.[4] At age 20, the chance is 1 in 1,441; at age 30, it is 1 in 959; at age 40, it is 1 in 84; and at age 50 it is 1 in 44.[4] Although the probability increases with maternal age, 70% of children with Down syndrome are born to women 35 years of age and younger, because younger people have more children.[4] The father's older age is also a risk factor in women older than 35, but not in women younger than 35, and may partly explain the increase in risk as women age.[134]

Lionel Penrose



Born	Lionel Sharples Penrose 11 June 1898[1] London, UK[3]
Died	12 May 1972 (aged 73) London, UK
Alma mater	St John's College, Cambridge University of Vienna King's College London
Known for	Penrose triangle Penrose method Penrose stairs[4] Penrose's Law[5][6] Penrose square root law Penrose–Banzhaf index
Spouse	Margaret Leathes (m. 1928)
Children	Oliver Penrose Roger Penrose Jonathan Penrose Shirley Hodgson
Awards	Fellow of the Royal Society[1] Lasker Award[2] James Spence Medal 1964.
	Scientific career
Fields	Pediatrics, Psychiatry, Genetics
Institutions	University of Cambridge University College London

Journal of Genetics, 1933

Journal of Genetics, 1933

THE RELATIVE EFFECTS OF PATERNAL AND MATERNAL AGE IN DOWN'S SYNDROME

BY L.S. PENROSE, M.D.*

In the human species ages of the parents are so closely correlated that it is difficult to separate the effects of the two elements. Hitherto attempts maternal age is of more aetiological importance than the paternal age. For example, a serious attempt to solve the problem was made in 1927 by Van der Scheer[2], who compared the relative percentages of 316 Down's syndrome children born at various maternal ages with the percentages of normal children born to mothers of equivalent ages in a very large series of families gathered from the general population. The resulting ratios showed a very marked increase in the incidence of affected childen as the age of the mother increased, and also a similar, though not quite so marked, rise of the incidence with increasing paternal age.

In studying the effects of parental age on certain characteristics in inbred stocks of guinea-pigs,

In studying the effects of parental age on certain characteristics in inbred stocks of guinea-pigs, Wright* was able to demonstrate that the age of the dam [mother] was an essential factor in the production of polydactyly [extra fingers].

In studying the effects of parental age on certain characteristics in inbred stocks of guinea-pigs, Wright* was able to demonstrate that the age of the dam [mother] was an essential factor in the production of polydactyly [extra fingers].

- The correlation of age of dam and proportion of polydactylous offspring was -0.370 $\pm$ 0.018.

In studying the effects of parental age on certain characteristics in inbred stocks of guinea-pigs, Wright* was able to demonstrate that the age of the dam [mother] was an essential factor in the production of polydactyly [extra fingers].

- The correlation of age of dam and proportion of polydactylous offspring was -0.370 $\pm$ 0.018.
- the use of the **technique of partial correlation** enabled him to show that the effect of the age of the sire [father] was negligible, after eliminating the effect of the age of the dam [mother].

*Wright, S. Effect of age of parents upon characteristics of the Guinea-pig. 1926

In studying the effects of parental age on certain characteristics in inbred stocks of guinea-pigs, Wright* was able to demonstrate that the age of the dam [mother] was an essential factor in the production of polydactyly [extra fingers].

- The correlation of age of dam and proportion of polydactylous offspring was -0.370 $\pm$ 0.018.
- the use of the **technique of partial correlation** enabled him to show that the effect of the age of the sire [father] was negligible, after eliminating the effect of the age of the dam [mother].

*Wright, S. Effect of age of parents upon characteristics of the Guinea-pig. 1926

BACK THEN, the use of a (Pearson) CORRELATION between a **binary (0/1) Y** and a **quantitative X** was COMMON.

Material

Material

The present writer has attempted a **similar treatment** of 150 families of the human species containing **Down's syndrome among the children**.

Material

The present writer has attempted a **similar treatment** of 150 families of the human species containing **Down's syndrome among the children**.

Every family included was visited personally and, among other things, the ages of the parents at the birth of each child was carefully recorded: miscarriages and all individuals in whom a diagnosis of normality or Down's syndrome could not be made with certainty were excluded.

No obvious disparity was observed between the ages of the parental pairs, which were distributed in a manner resembling that found by pooling all married couples in the general population.

150 Families

N = Normal; D = Down's

Maternal Age (q)

Paternal age (p)	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	N	D
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
19	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
20	.	.	2	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
21	.	.	1	2	2	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.
22	.	.	.	1	4	3	.	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	2
23	.	.	2	1	1	2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	2
24	.	.	.	.	2	3	4	5	2	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	16	2
25	.	.	.	1	2	.	4	6	7	4	4	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	27	3
26	.	.	.	3	2	2	3	2	3	5	1	3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	1
27	.	.	.	.	.	1	3	3	2	5	7	1	1	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	3
28	.	.	.	.	.	1	3	5	.	4	8	2	3	4	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	28	6
29	1	.	.	.	.	.	2	1	3	1	3	3	6	2	1	2	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	25	2
30	.	1	.	.	.	.	.	2	3	4	2	5	4	5	1	1	1	2	1	.	.	.	.	.	1	.	.	.	.	.	.	31	2
31	.	.	1	.	1	.	1	.	4	3	4	3	3	6	7	3	4	1	2	.	.	.	.	.	.	.	.	.	.	.	.	39	4
32	.	.	.	1	.	.	.	1	.	1	1	2	1	5	4	6	2	2	1	.	.	.	1	.	.	.	.	.	.	.	.	26	2
33	.	.	.	.	1	.	1	1	.	.	2	3	5	4	3	1	5	1	1	2	.	1	.	.	.	.	.	.	.	.	.	28	3
34	.	.	.	.	1	.	.	.	.	1	.	.	.	5	.	5	8	5	2	1	1	2	.	.	1	.	.	.	.	.	.	30	2
35	.	.	.	.	.	.	.	.	.	1	.	.	1	1	3	3	4	6	10	4	3	3	1	.	.	.	.	.	.	.	.	33	7
36	.	.	.	.	.	.	1	2	.	.	2	.	.	.	6	2	3	3	4	5	1	3	.	2	.	.	.	.	.	.	.	32	2
37	.	.	.	.	.	.	.	1	.	.	.	.	1	.	1	2	6	2	5	2	9	2	2	1	1	1	.	.	.	.	.	27	9
38	.	.	.	.	.	.	.	.	1	1	.	.	2	.	.	1	2	2	2	4	5	4	1	2	2	.	.	.	.	.	.	21	8
39	.	.	.	.	.	.	.	1	.	2	1	.	.	.	1	.	2	4	6	2	2	2	7	1	.	1	1	.	.	.	.	25	8
40	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	1	.	1	3	1	.	2	3	4	3	2	1	1	.	.	.	17	7
41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	2	.	.	3	4	4	1	1	.	.	.	.	10	8
42	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	2	.	.	2	6	2	3	3	2	.	1	.	1	.	9	15
43	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	1	3	1	3	3	4	3	2	.	.	.	11	11
44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.	1	2	3	5	1	3	1	2	.	.	.	11	9
45	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	2	1	.	1	1	2	.	1	1	1	.	3	3	3	2	.	15	7
46	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1	1	1	2	1	1	.	4	4
47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	2	.	.	.	.	.	1	5	.	.	2	1	.	7	6
48	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	1	.	.	2	1	1	1	1	.	.	2	1	6	6
49	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	1	.	2	4	1
50	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	1	1	.	.	.	.	4	.
51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	1	1	.	.	.	1	.	3	2
52	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	1	1	.	2	2
53	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	1	.	.	.	.	.	.	.	.	1	2
54	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	2	.	.	3	1
55	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1	1	1
56	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	1	1
57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1	.
58	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	2
59	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	1	.
N	1	2	5	8	13	11	22	31	25	31	31	22	30	36	32	32	35	27	35	21	21	22	21	16	10	9	6	5	8	4	3	573	
D	.	.	2	1	3	2	3	1	.	3	4	2	1	.	2	2	6	7	9	5	9	11	10	13	15	16	8	7	4	5	1		154

Correlations obtained from data summarised in Table I:

D = Indicator (0/1) of Down's syndrome

Correlations obtained from data summarised in Table I:

D = Indicator (0/1) of Down's syndrome

r(D, Father'sAge)
+0.294 $\pm$ 0.034

r(D, Mother'sAge)
+0.362 $\pm$ 0.032

Correlations obtained from data summarised in Table I:

D = Indicator (0/1) of Down's syndrome

r(D, Father'sAge)
+0.294 $\pm$ 0.034

r(D, Mother'sAge)
+0.362 $\pm$ 0.032

r(Father'sAge,Mother'sAge)
+0.829 $\pm$ 0.012

Correlations obtained from data summarised in Table I:

D = Indicator (0/1) of Down's syndrome

r(D, Father'sAge)
+0.294 $\pm$ 0.034

r(D, Mother'sAge)
+0.362 $\pm$ 0.032

r(Father'sAge,Mother'sAge)
+0.829 $\pm$ 0.012

PARTIAL Correlations (r_p)

Correlations obtained from data summarised in Table I:

D = Indicator (0/1) of Down's syndrome

r(D, Father'sAge)
+0.294 $\pm$ 0.034

r(D, Mother'sAge)
+0.362 $\pm$ 0.032

r(Father'sAge,Mother'sAge)
+0.829 $\pm$ 0.012

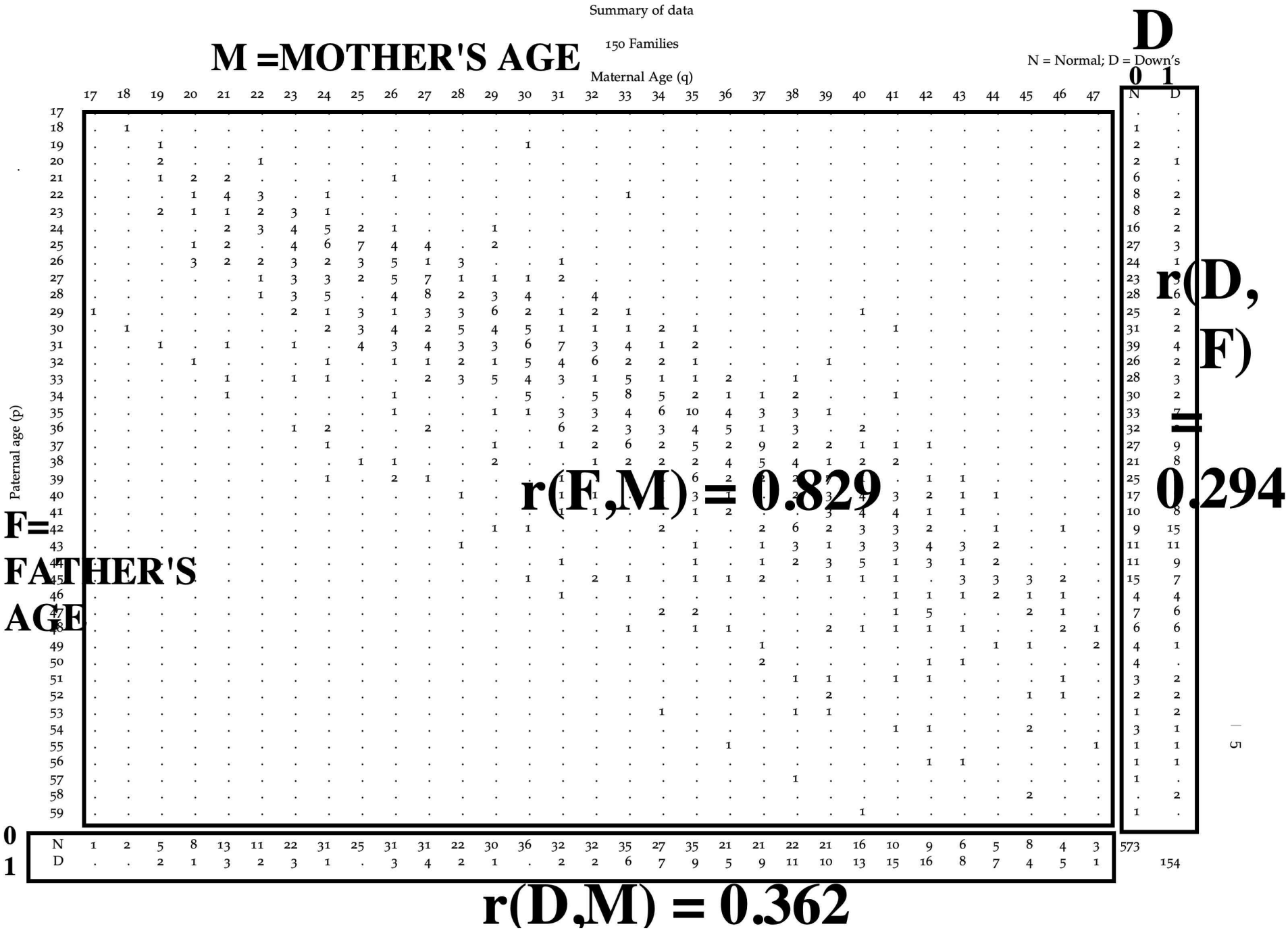
PARTIAL Correlations (r_p)

r_p(D, Father'sAge)
eliminating Mother's Age

-0.011 $\pm$ 0.04

r_p(D, Mother'sAge)
eliminating Father's Age

+0.221 $\pm$ 0.04


Summary of data
150 Families
M =MOTHER'S AGE
Maternal Age (q)
D
N = Normal; D = Down's
0 1
N D
Paternal age (p)
F= FATHER'S AGE
r(F,M) = 0.829
r(D, F) = 0.294
r(D,M) = 0.362
0
1
N 1 2 5 8 13 11 22 31 25 31 31 22 30 36 32 32 35 27 35 21 21 22 21 16 10 9 6 5 8 4 3 573
D . . 2 1 3 2 3 1 . 3 4 2 1 . 2 2 6 7 9 5 9 11 10 13 15 16 8 7 4 5 1 154

“The partial correlation technique is open to certain objections"

“The partial correlation technique is open to certain objections"

“ ... partly on account of the hypothetical nature of the [0/1] variable Down’s syndrome - Normal.

“The partial correlation technique is open to certain objections"

“ ... partly on account of the hypothetical nature of the [0/1] variable Down’s syndrome - Normal.

It is not, however, necessary to use this technique to demonstrate the various points,

“The partial correlation technique is open to certain objections"

“ ... partly on account of the hypothetical nature of the [0/1] variable Down’s syndrome - Normal.

It is not, however, necessary to use this technique to demonstrate the various points,

and the following method is at once clearer and more precise.”

Observed (& ‘expected’) mean parental ages of Down’s *vs* Normal children

Observed (& 'expected') mean parental ages of Down's *vs* Normal children

	Father		Mother	
	Obs.	Exp.	Obs.	Exp.
Down's	39.383	39.471	37.253	35.712
Normal	33.830	33.803	31.249	31.680
Difference	5.553	5.668	6.004	4.032

Observed (& 'expected') mean parental ages of Down's *vs* Normal children

	Father			Mother	
	Obs.		Exp.	Obs.	Exp.
Down's	39.383		39.471	37.253	35.712
Normal	33.830		33.803	31.249	31.680
Difference	5.553		5.668	6.004	4.032
		-0.115		**+1.972**	
		(**SE 0.392**)		(**SE 0.341**)	

Observed (& 'expected') mean parental ages of Down's *vs* Normal children

	Father			Mother	
	Obs.		Exp.	Obs.	Exp.
Down's	39.383		39.471	37.253	35.712
Normal	33.830		33.803	31.249	31.680
Difference	5.553		5.668	6.004	4.032
		-0.115		**+1.972**	
		(**SE 0.392**)		(**SE 0.341**)	

$$\widehat{AgeOfFather} \Leftarrow 4.304 + 0.944 \times AgeOfMother$$

$$7.120 + 0.726 \times AgeOfFather \Rightarrow \widehat{AgeOfMother}$$

Observed (& 'expected') mean parental ages of Down's *vs* Normal children

	Father			Mother	
	Obs.		Exp.	Obs.	Exp.
Down's	39.383		39.471	37.253	35.712
Normal	33.830		33.803	31.249	31.680
Difference	5.553		5.668	6.004	4.032
		-0.115		**+1.972**	
		(**SE 0.392**)		(**SE 0.341**)	

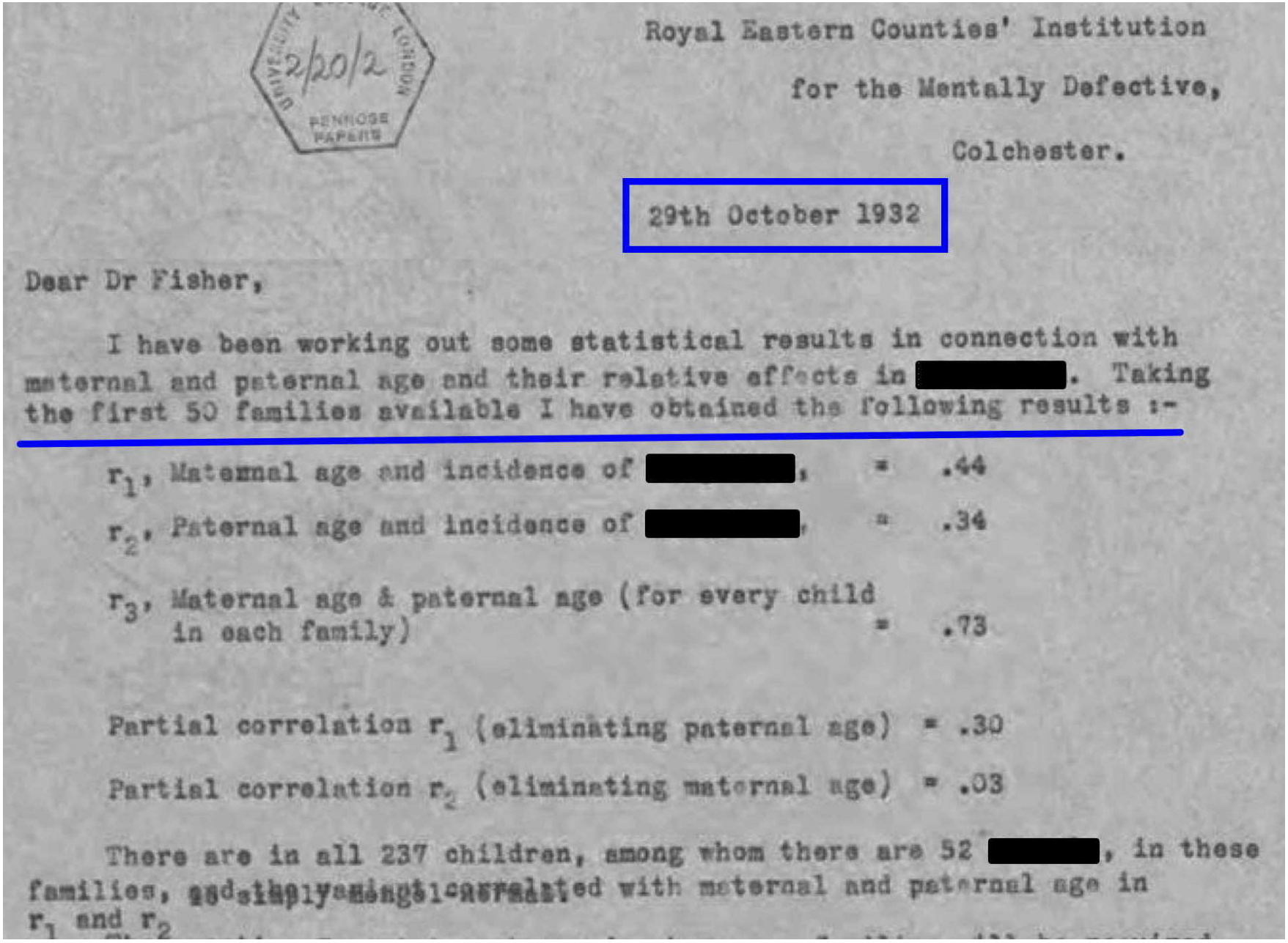
$$\widehat{AgeOfFather} \Leftarrow 4.304 + 0.944 \times AgeOfMother$$

$$7.120 + 0.726 \times AgeOfFather \Rightarrow \widehat{AgeOfMother}$$

"There can be little doubt, judging from these results, which confirm those obtained by the partial correlation technique, that the father's age is an insignificant factor in the aetiology of Down's syndrome, the emphasis being entirely on the age of the mother."

Penrose consults Fisher: 1932

Penrose consults Fisher: 1932



UNIVERSITY COLLEGE LONDON
2/20/2
PENROSE PAPERS

Royal Eastern Counties' Institution
for the Mentally Defective,
Colchester.

29th October 1932

Dear Dr Fisher,

I have been working out some statistical results in connection with maternal and paternal age and their relative effects in [redacted]. Taking the first 50 families available I have obtained the following results :-

r_1, Maternal age and incidence of [redacted], = .44

r_2, Paternal age and incidence of [redacted], = .34

r_3, Maternal age & paternal age (for every child in each family) = .73

Partial correlation r_1 (eliminating paternal age) = .30

Partial correlation r_2 (eliminating maternal age) = .03

There are in all 237 children, among whom there are 52 [redacted], in these families, and highly significantly correlated with maternal and paternal age in r_1 and r_2

There are in all 237 children, among whom there are 52 [redacted], in these families, and they are singly correlated with maternal and paternal age in r_1 and r_2.

The question I want to ask you is, how many families will be required to give significance to these results, which appear to show that maternal age is of importance but paternal age insignificant ? I have available nearly 200 families and have every reason to suppose that similar figures will be obtained from the whole group to those I have found in the first 50 families. The question of significance is rather troublesome because I am not sure what the effect is of having several of the children in the same family and am doubtful whether the ordinary methods of finding the standard deviations of these coefficients are applicable here. If you think that about 180 families would give a significant result, it seems to me that it is an elegant method of demonstrating what has been rather a vexed question for some time.

Please excuse my bothering you in this way, but technical points of this kind frequently arise in my work and it is a very great help to have your advice on such matters.

Yours sincerely

31 October 1932.

Dr. L. Penrose,
Royal Eastern Counties' Institution,
Colchester.

Dear Dr. Penrose:

I do not understand what is the variate you call incidence of [redacted]. As far as I can see you have two observed two-way distributions of Mothers' age (x) and Fathers' age (y), one for [redacted] and one for normals.

These supply a common value (based on pooled products) of the regression of Fathers' age on Mothers' age, i.e. of y on x:

$$Y = bx + c.$$

You wish to compare $\bar{x}$ for [redacted] with $\bar{x}$ for normals, and this is a straight t test.

Equally you can compare $\bar{y} - b\bar{x}$ for [redacted] v. normals, with a known standard error.

2014

Regression-type models for ‘imaginary’ response variates: 1933-

[* Fisher’s derogatory term for BINARY variates used in a Pearson correlation]

Regression-type models for 'imaginary' response variates: 1933-

[* Fisher's derogatory term for BINARY variates used in a Pearson correlation]

1936	Fisher	The Use of Multiple Measurements in Taxonomic Problems. (LDA) Ann. Eugenics.
1952	Duncan, Rhodes	Multiple [Probit] Regression with a Quantal Response (Meeting Abstract)
1955	Berkson	Max. likelihood and min. χ^2 estimates of the logistic function (JASA)
1958	Cox	The regression analysis of binary sequences (JRSS-B)
1962	Cornfield	Joint dependence of risk of CHD on ... : a discriminant function analysis (Fed. Proc)
1966	Cox	Some procedures connected with the logistic qualitative response curve. (Essays)
1967	Cornfield, Kannel	Multivariate Analysis of Risk of Coronary Heart Disease in Framingham (JChronicDis)
1967	Walker, Duncan	Estimation of Probability of an Event as Function of Several Independent Variables (B'ka)
1972	Nelder, Wedderburn	Generalized Linear Models (JRSS-A)
1973	McFadden	Conditional logit analysis of qualitative choice behavior (Book Chapter, Economics)

Wed 2014-12-31 2:37 PM

Dear Dr Hanley,

Thank you for your interesting message.

Wed 2014-12-31 2:37 PM

Dear Dr Hanley,

Thank you for your interesting message.

[... does not have raw material; hints at where I might look for it]

Wed 2014-12-31 2:37 PM

Dear Dr Hanley,

Thank you for your interesting message.

[... does not have raw material; hints at where I might look for it]

I have also written a paper commenting on this work of my father's. When I wrote it* I was not aware of my father's correspondence with R A Fisher, and my paper may have other flaws as well — it was not refereed. It was published as
A beautiful method of analysis in "Fifty years of human genetics: A Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser", *2007.

Wed 2014-12-31 2:37 PM

Dear Dr Hanley,

Thank you for your interesting message.

[... does not have raw material; hints at where I might look for it]

I have also written a paper commenting on this work of my father's. When I wrote it* I was not aware of my father's correspondence with R A Fisher, and my paper may have other flaws as well — it was not refereed. It was published as
A beautiful method of analysis in "Fifty years of human genetics: A Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser", *2007.

Your "sudoku" problem appears to have multiple solutions in general. For example, suppose the **matrix** were

		N	D
5	**5**	5	5
5	**5**	5	5

	N	D
N	5	5
D	5	5

then the 6 solutions are

5,0	0,5	4,1	1,4	. . .	0,5	5,0
0,5	5,0	1,4	4,1	. . .	5,0	0,5

With best wishes for the new year

Oliver Penrose

'Statistical Sudoku' Challenge: splitting the Penrose frequency table

This table is from the 1933 article *'The relative effects of paternal and maternal age' in Down's syndrome* by the geneticist Lionel Penrose.

The row totals at the right of the table contain the paternal-age-specific frequencies for the 'Normal' (N) and Down's children (D) separately. Likewise, the columns totals at the bottom contain separate maternal-age-specific frequencies for the 'Normal' and Down's children.

However, the frequencies in the body of the table are the **sums** of the frequencies of Normal and Down's cases.

The **CHALLENGE** is to split each row-column-specific sum into the frequency of N and the frequency of D cases.

Screenshot

Summary of data

150 Families

N = Normal; D = Down's

Paternal age (p) \ Maternal Age (q)	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	N	D
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
19	.	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
20	.	.	2	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
21	.	.	1	2	2	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.
22	.	.	.	1	4	3	.	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	2
23	.	.	2	1	1	2	3	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	2
24	.	.	.	.	2	3	4	5	2	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	16	2
25	.	.	.	1	2	.	4	6	7	4	4	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	27	3
26	.	.	.	3	2	2	3	2	3	5	1	3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	1
27	.	.	.	.	.	1	3	3	2	5	7	1	1	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	3
28	.	.	.	.	.	1	3	5	.	4	8	2	3	4	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	28	6
29	1	.	.	.	.	.	2	1	3	1	3	3	6	2	1	2	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	25	2
30	.	1	.	.	.	.	.	2	3	4	2	5	4	5	1	1	1	2	1	.	.	.	.	.	1	.	.	.	.	.	.	31	2
31	.	.	1	.	1	.	1	.	4	3	4	3	3	6	7	3	4	1	2	.	.	.	.	.	.	.	.	.	.	.	.	39	4
32	.	.	.	1	.	.	.	1	.	1	1	2	1	5	4	6	2	2	1	.	.	.	1	.	.	.	.	.	.	.	.	26	2
33	.	.	.	.	1	.	1	1	.	.	2	3	5	4	3	1	5	1	1	2	.	1	.	.	.	.	.	.	.	.	.	28	3
34	.	.	.	.	1	.	.	.	.	1	.	.	.	5	.	5	8	5	2	1	1	2	.	.	1	.	.	.	.	.	.	30	2
35	.	.	.	.	.	.	.	.	.	1	.	.	1	1	3	3	4	6	10	4	3	3	1	.	.	.	.	.	.	.	.	33	7
36	.	.	.	.	.	.	1	2	.	.	2	.	.	.	6	2	3	3	4	5	1	3	.	2	.	.	.	.	.	.	.	32	2
37	.	.	.	.	.	.	.	1	.	.	.	.	1	.	1	2	6	2	5	2	9	2	2	1	1	1	.	.	.	.	.	27	9
38	.	.	.	.	.	.	.	.	1	1	.	.	2	.	.	1	2	2	2	4	5	4	1	2	2	.	.	.	.	.	.	21	8
39	.	.	.	.	.	.	.	1	.	2	1	.	.	.	1	.	2	4	6	2	2	2	7	1	.	1	1	.	.	.	.	25	8
40	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	1	.	1	3	1	.	2	3	4	3	2	1	1	.	.	.	17	7
41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	2	.	.	3	4	4	1	1	.	.	.	.	10	8
42	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	2	.	.	2	6	2	3	3	2	.	1	.	1	.	9	15
43	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	1	3	1	3	3	4	3	2	.	.	.	11	11
44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	.	1	2	3	5	1	3	1	2	.	.	.	11	9

https://jhanley.biostat.mcgill.ca/StatisticalSudoku/

N	1	2	5	8	13	11	22	31	25	31	31	22	30	36	32	32	35	27	35	21	21	22	21	16	10	9	6	5	8	4	3	573	
D	.	.	2	1	3	2	3	1	.	3	4	2	1	.	2	2	6	7	9	5	9	11	10	13	15	16	8	7	4	5	1			154

It is easy to split a few of the small (N + D) counts into their separate N and D components. e.g. N[18,18] = 1; D[18,18] = 0.
But for larger (N + D) counts closer to the centre of the table, the task of splitting them into N and D becomes progressively more difficult.

Background

Down syndrome or Down's syndrome, also known as trisomy 21, is now known to be a genetic disorder caused by the presence of all or part of a third copy of chromosome 21. It is typically associated with physical growth delays, characteristic facial features, and mild to moderate intellectual disability.

Screenshot

The relative effects of paternal and maternal age in Down's syndrome by Lionel Penrose, was published in the Journal of Genetics, May 1933, Volume 27, Issue 2, pp 219-224. This was before its genetic nature was known. Penrose was one of the first to establish that its frequency rate was a stronger function of maternal than paternal age, and (in a separate article in 1934) of maternal age than birth order

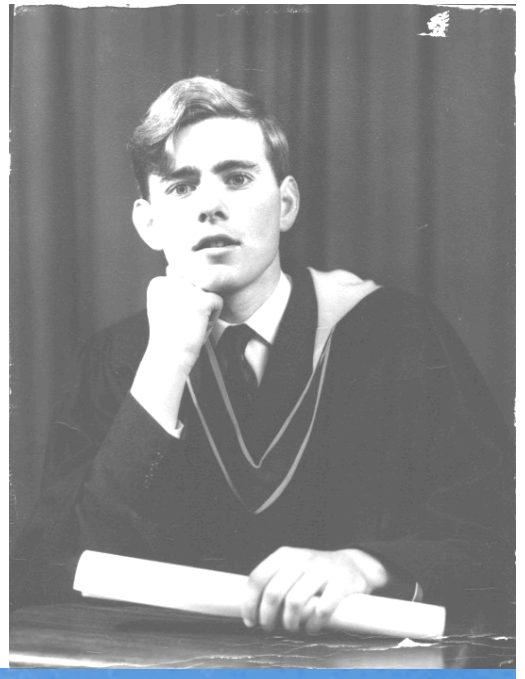
JH became aware of Penrose's work through the Wellcome Library's recent digitization of The Lionel Penrose papers. These papers also contain extensive correspondence with statistician Ronald Fisher, a pioneer in statistical genetics.

Lionel Penrose was the father of mathematician **Roger Penrose**, chess Grandmaster **Jonathan Penrose**, mathematician **Oliver Penrose** (who graciously provided JH with this reconsideration of his father's 1933 work. *[It appears as the chapter 'A beautiful method of analysis': O. Penrose, pp 443-451 in "Fifty years of human genetics: A Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser", edited by Oliver Mayo and Carolyn Leach, ISBN 9781862547537, Wakefield Press, Adelaide, 2007]*), and geneticist **Shirley Victoria Hodgson** (who graciously sent JH a copy of her father's book "Clinical and Genetic Study of 1280 Cases of Mental Defect: Colchester Survey").

James Hanley,
Dept of Epidemiology, Biostatistics and Occupational Health,
McGill University

jhanley.biostat.mcgill.ca

UNIVERSITY COLLEGE CORK
Coláiste na hOllscoile Corcaigh



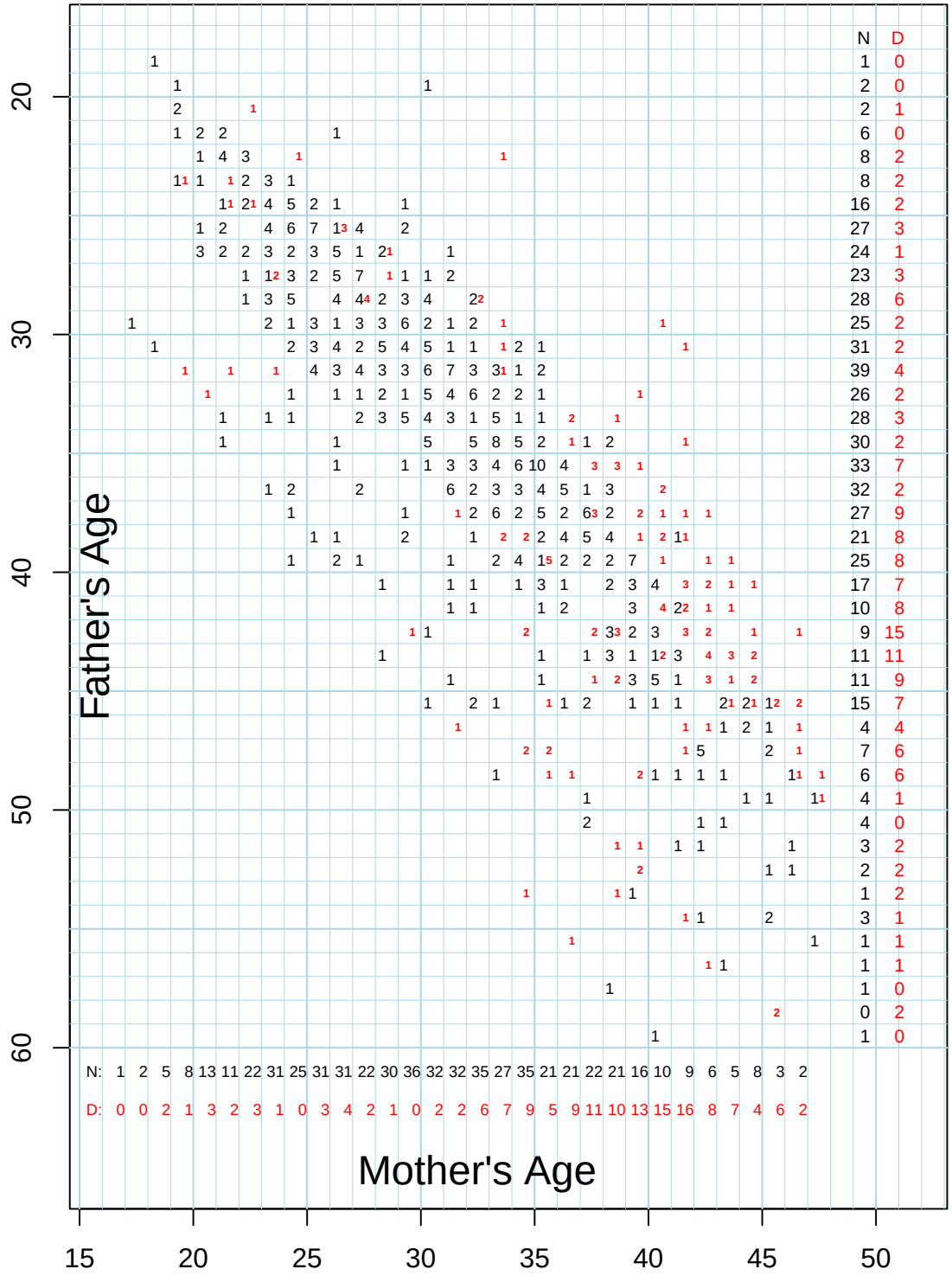
BSc (Mathematics/Statistics) 1968
MSc (Mathematics/Statistics) 1969


STATISTICAL SUDOKU
2023
Supratik Roy
19:21
57:18

SUDOKU !

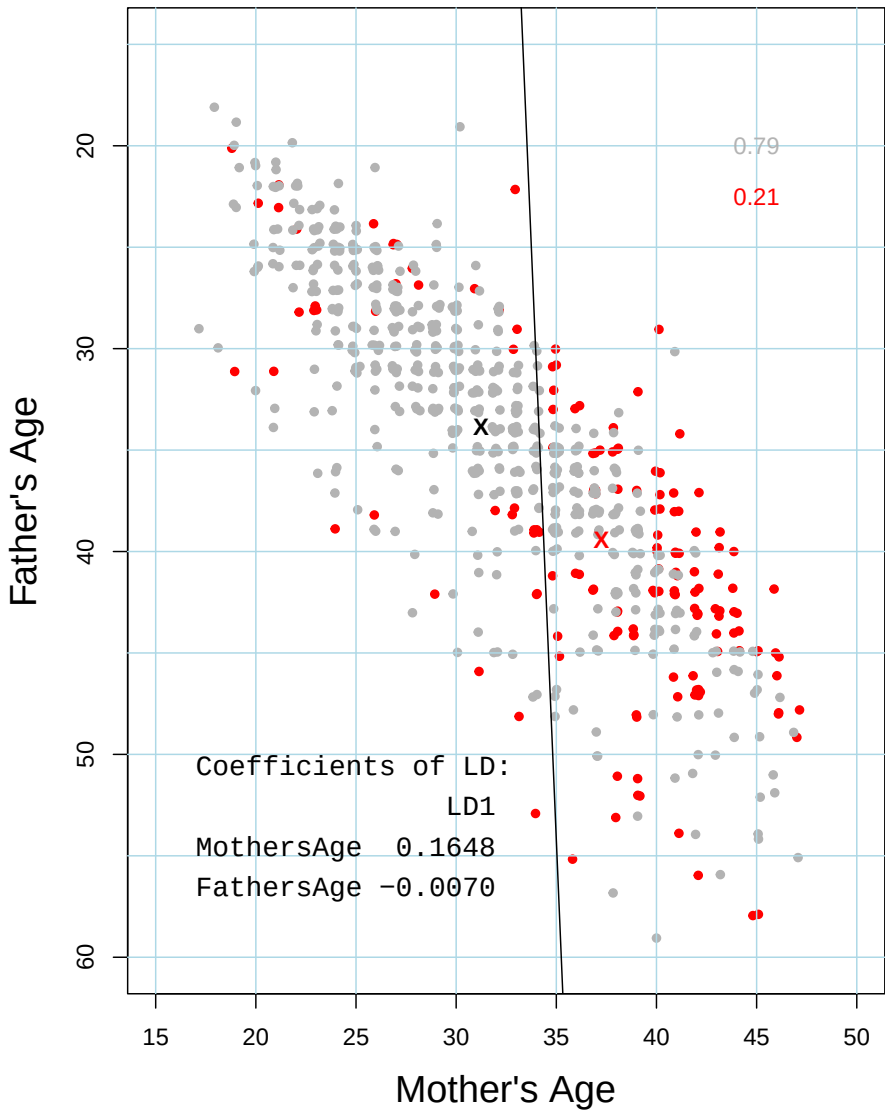
SUDOKU !

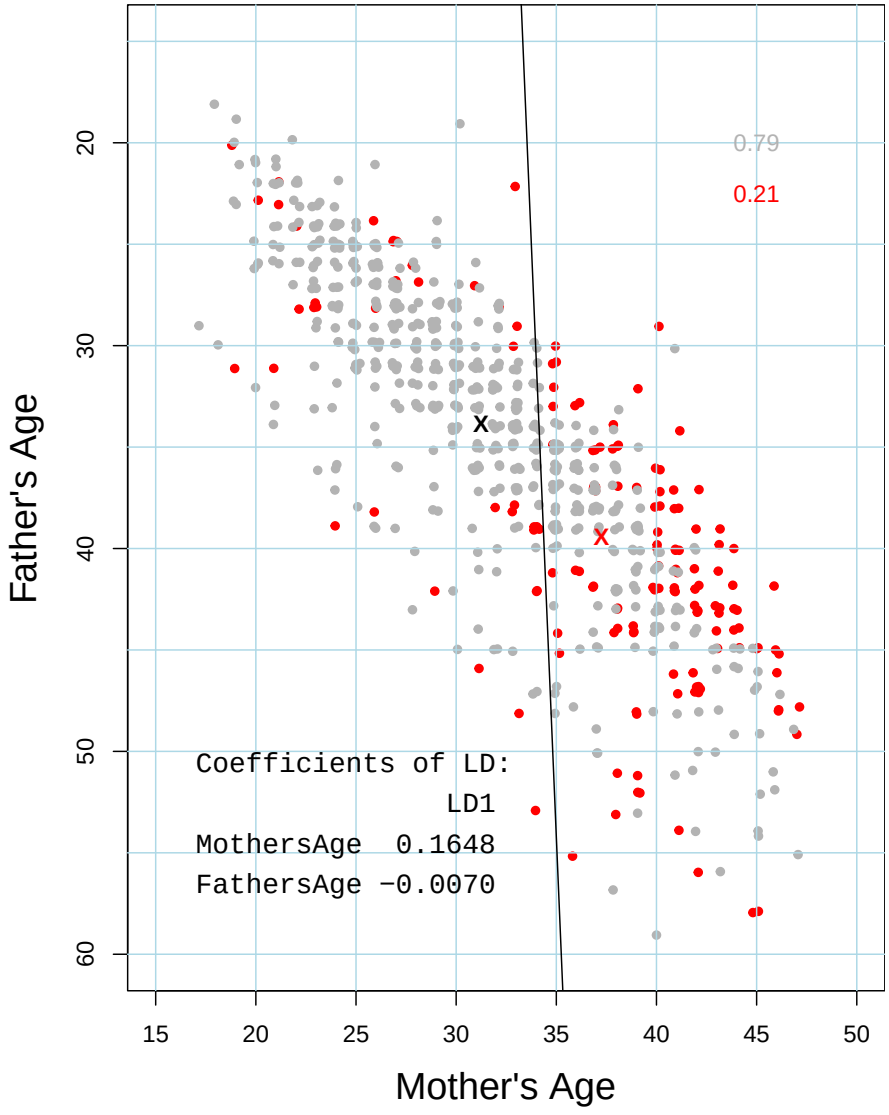
For each of the 325 cells,
shown are the number of children

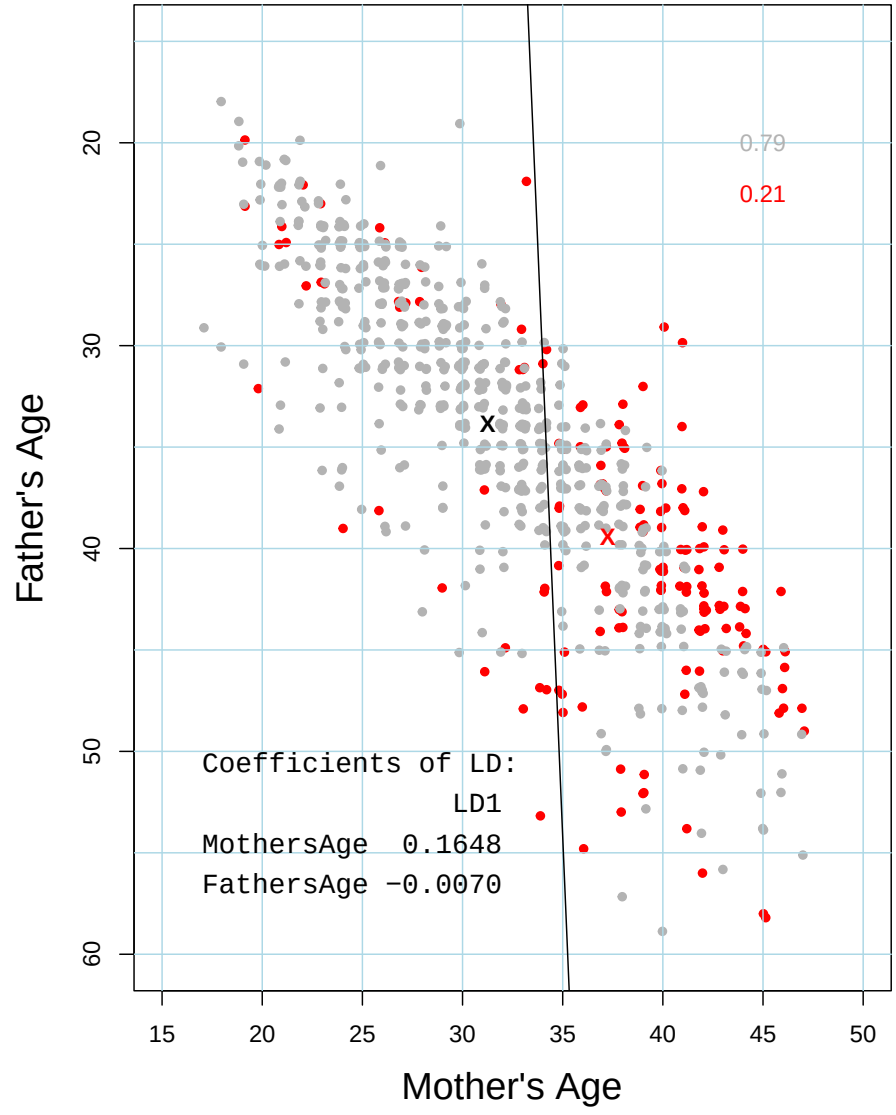

Father's Age
Mother's Age
20
30
40
50
60
15
20
25
30
35
40
45
50
N D
N: 1 2 5 8 13 11 22 31 25 31 31 22 30 36 32 32 35 27 35 21 21 22 21 16 10 9 6 5 8 3 2
D: 0 0 2 1 3 2 3 1 0 3 4 2 1 0 2 2 6 7 9 5 9 11 10 13 15 16 8 7 4 6 2

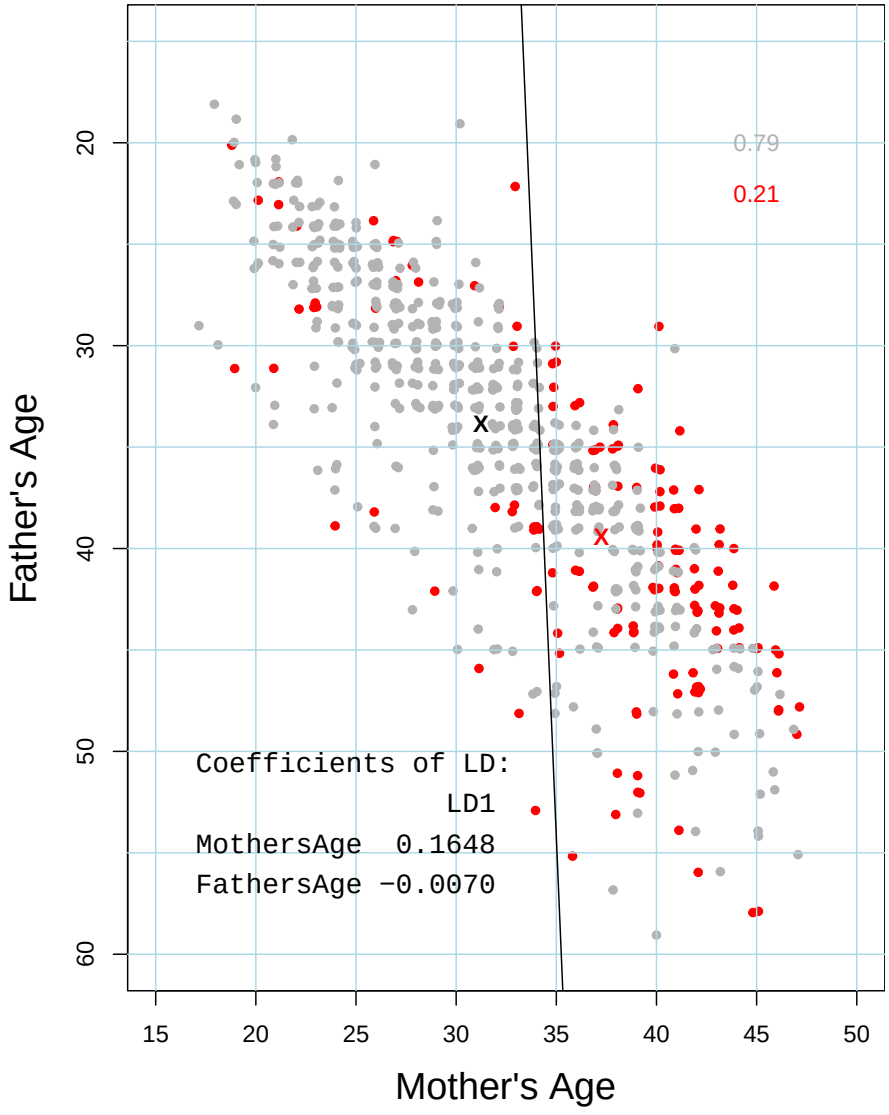
JH's first analyses of 3-D datasets

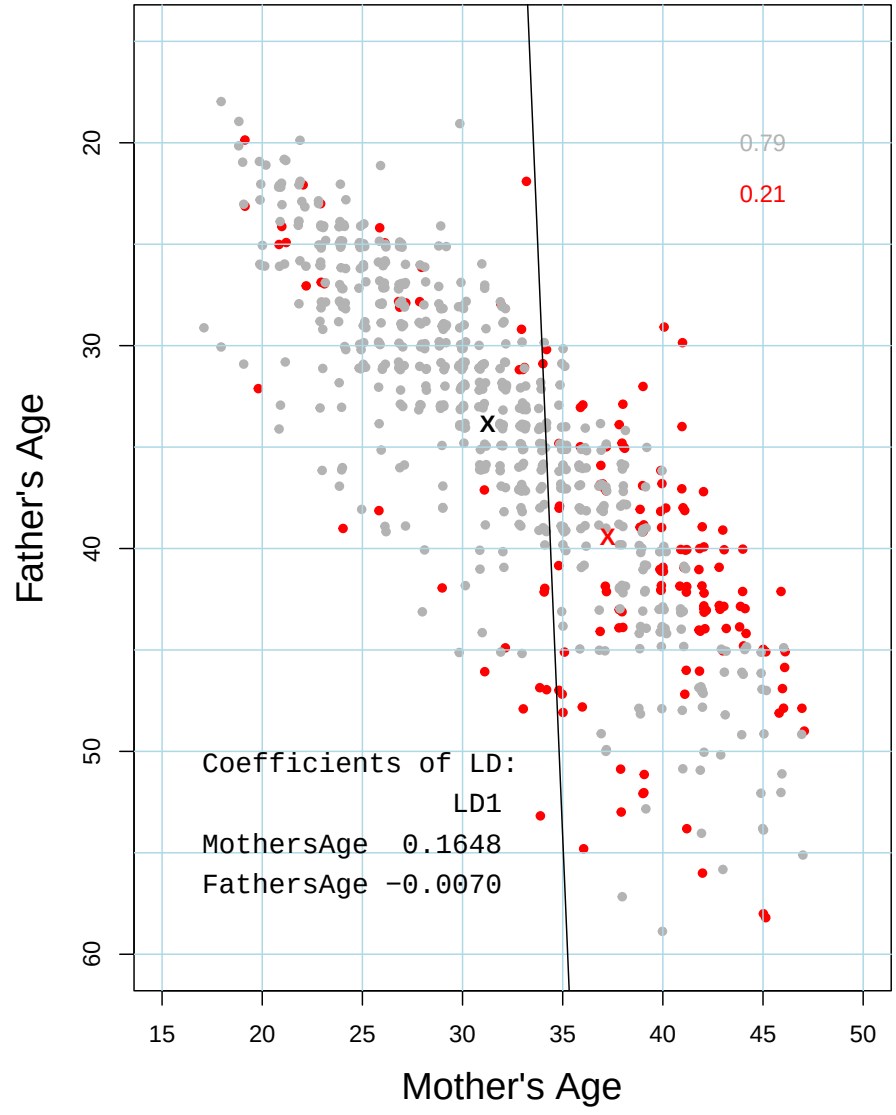
- Linear Discriminant Analysis
- Logistic regression
- And the surprises he got!


Father's Age
Mother's Age
20
30
40
50
60
15
20
25
30
35
40
45
50
0.79
0.21
Coefficients of LD:
LD1
MothersAge 0.1648
FathersAge −0.0070


0.79
0.21
Coefficients of LD:
LD1
MothersAge 0.1648
FathersAge −0.0070
Father's Age
Mother's Age
15
20
25
30
35
40
45
50
20
30
40
50
60


0.79
0.21
Coefficients of LD:
LD1
MothersAge 0.1648
FathersAge −0.0070
Father's Age
Mother's Age
15
20
25
30
35
40
45
50
20
30
40
50
60





LDA: Same fit no matter the solution !

Logistic regression

```
glm(formula = Downs ~  ... , family = binomial)

Coefficients:
             Estimate Std.Error z.value Pr(>|z|)

(Intercept) -4.796     0.489   -9.80   < 2e-16 ***
FathersAge   0.095     0.012    7.51 5.62e-14 ***


(Intercept) -6.526     0.610 -10.69    <2e-16 ***
MothersAge   0.151     0.016    9.04    <2e-16 ***


(Intercept) -6.491     0.614 -10.57   < 2e-16 ***
MothersAge   0.162     0.027    5.94  2.8e-09 ***
FathersAge  -0.011     0.022   -0.51     0.604
```

Logistic regression

```
glm(formula = Downs ~  ... , family = binomial)

Coefficients:
             Estimate Std.Error z.value Pr(>|z|)

(Intercept) -4.796     0.489   -9.80   < 2e-16 ***
FathersAge   0.095     0.012    7.51 5.62e-14 ***


(Intercept) -6.526     0.610 -10.69    <2e-16 ***
MothersAge   0.151     0.016    9.04    <2e-16 ***


(Intercept) -6.491     0.614 -10.57   < 2e-16 ***
MothersAge   0.162     0.027    5.94  2.8e-09 ***
FathersAge  -0.011     0.022   -0.51     0.604
```

Same fit no matter the solution !

The takeaway

The takeaway

- All feasible solutions give the same statistics that Penrose reports in his paper.

The takeaway

- All feasible solutions give the same statistics that Penrose reports in his paper.
- i.e., these statistics are independent of what Penrose's 3-D dataset *might have been*.

The takeaway

- All feasible solutions give the same statistics that Penrose reports in his paper.
- i.e., these statistics are independent of what Penrose's 3-D dataset *might have been*.
- The three 2-D marginal tables that Penrose published were sufficient to derive his statistics.

The takeaway

- All feasible solutions give the same statistics that Penrose reports in his paper.
- i.e., these statistics are independent of what Penrose's 3-D dataset *might have been*.
- The three 2-D marginal tables that Penrose published were sufficient to derive his statistics.
- Recall Fisher on sufficiency and what use can be made of the remainder of the data.

The takeaway

- All feasible solutions give the same statistics that Penrose reports in his paper.
- i.e., these statistics are independent of what Penrose's 3-D dataset *might have been*.
- The three 2-D marginal tables that Penrose published were sufficient to derive his statistics.
- Recall Fisher on sufficiency and what use can be made of the remainder of the data.
- The 'Goodness of Fit' statistics DO vary from solution to solution; Penrose did not supply any G.o.F measure for his dataset. See webpage for SR's explorations of 11 solution sets.

What summary statistics are needed to fit a logistic regression?

What summary statistics are needed to fit a logistic regression?

Example with 2 predictors, x (age of mother) and z (age of father)

$$P[D=1] = \frac{e^{q_1(x)+q_2(z)}}{1+e^{q_1(x)+q_2(z)}}, \quad P[D=0] = \frac{1}{1+e^{q_1(x)+q_2(z)}}.$$

$$L\left(\{d_{ijk}, x_i, z_j\}\right) = \prod_{i,j,k} \left(\frac{e^{q_1(x_i)+q_2(z_j)}}{1+e^{q_1(x_i)+q_2(z_j)}}\right)^{d_{ijk}} \left(\frac{1}{1+e^{q_1(x_i)+q_2(z_j)}}\right)^{1-d_{ijk}}.$$

$$logL = \sum_{i,j}\left[\sum_k d_{ijk} q_1(x_i) + \sum_k d_{ijk} q_2(z_j) - \sum_k \ln\left(1+e^{q_1(x_i)+q_2(z_j)}\right)\right]$$

$$= \sum_{i,j}\left[d_{ij\bullet} q_1(x_i) + d_{ij\bullet} q_2(z_j) - f_{ij}\sum_k \ln\left(1+e^{q_1(x_i)+q_2(z_j)}\right)\right]$$

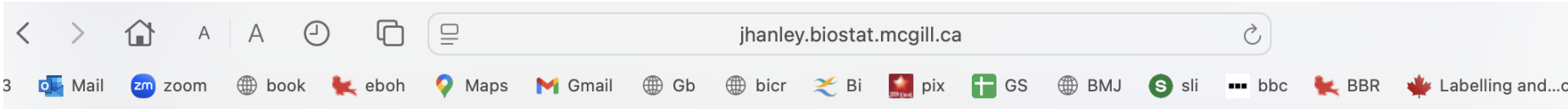
where f_{ij} is the total number of observations in the (i,j)-th cell, and q_1 and q_2 are unknown functions of x and z.

$$= \underbrace{\sum_i d_{i\bullet\bullet} q_1(x_i)} + \underbrace{\sum_j d_{\bullet j\bullet} q_2(y_j)} - \sum_{i,j} f_{ij} \sum_k \ln\left(1 + e^{q_1(x_i)+q_2(y_j)}\right).$$

- For the Penrose data set, we have the marginals $d_{i\bullet\bullet}$, $d_{\bullet j\bullet}$ and f_{ij} are the sum of Downs and non-Downs.
- Therefore the log likelihood remains invariant over all possible distributions of Downs as long as the marginals and the distribution of sums remain the same.
- Therefore maximum likelihood estimates for usual logistic regression or even spline logistic regression will be unique and independent of the actual distribution of Downs.
- However, the **fit** of each solution to the estimated likelihood will **differ**. From a model based approach we can consider the distribution of Downs that gives the best fit to a given model.

In *simplest* case, where $q_1(a) = \beta_1 a$ and $q_2(a) = \beta_2 a$, the 3 sufficient statistics for $\{\beta_0, \beta_1, \beta_2\}$ are the number of **cases**: 154; the sum of the ages of their 154 mothers: 5736 years; and their 154 fathers 6065 years.

EXPLORING SOLUTION SETS / co-au. Roy



contain extensive correspondence with statistician Ronald Fisher, a pioneer in statistical genetics.

Lionel Penrose was the father of mathematician **Roger Penrose**, chess Grandmaster **Jonathan Penrose**, mathematician **Oliver Penrose** (who graciously provided JH with this reconsideration of his father's 1933 work. *[It appears as the chapter 'A beautiful method of analysis': O. Penrose, pp 443-451 in "Fifty years of human genetics: A Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser", edited by Oliver Mayo and Carolyn Leach, ISBN 9781862547537, Wakefield Press, Adelaide, 2007]*), and geneticist **Shirley Victoria Hodgson*** (who graciously sent JH a copy of her father's book "Clinical and Genetic Study of 1280 Cases of Mental Defect: Colchester Survey").

2023

Prob[Down Syndrome | Parents' Ages]:
Statistical Sudoku and Analyses of Penrose's Data (J. Genetics 1933)
2023.11.22 Presentation by J. Hanley And S. Roy, McGill Biostatistics Seminar Series
[.mp4 video : *Penrose's daughter Shirley, born in London, Ontario, appears from London, England in last few frames]

2025

R code to read in Penrose's data and generate `Sudoku' solution sets

Appendices to manuscript, now-accepted by CHANCE**. These were shared with reviewers

**The accepted article
"Lionel Penrose's statistical consultant: and lessons from the statistical 'sudoku' they left us"
is planned to be published in the next issue.

James Hanley,
Dept of Epidemiology, Biostatistics and Occupational Health,
McGill University

jhanley.biostat.mcgill.ca

Lionel Penrose's Statistical Consultant and Lessons from the Statistical "Sudoku" They Left Us

James A. Hanley and Supratik Roy

Lionel Penrose's 1933 study was one of the first to establish that "the father's age is not a significant factor in the risk of Down's syndrome, while the mother's age is to be regarded as very important." We identify the statistical consultant who helped Penrose improve on the prevailing regression methods for handling binary responses. Using a form of "statistical sudoku," we explain how to create complete data sets from the summary data Penrose worked with and shared, and we describe the

[...] 150 families of the human species containing Down's syndrome among the children. Every family included was visited personally and, among other things, the ages of the parents at the birth of each child was [sic] carefully recorded: miscarriages and all individuals in whom a diagnosis of normality or Down's syndrome could not be made with certainty were excluded. No obvious disparity was observed between the ages

the three (2-D) marginal distributions but stopped short of giving the full 3-D distribution. Penrose omitted the cell-specific distributions of N and D's.

Penrose's challenge was how to deal with the tricky statistical issue that today we call confounding. The strong correlation of the parent's ages means that when the father's age or the mother's age is considered on its own, the probability that a child is affected by Down's syndrome will seem to be strongly related to that parent's age.

CHANCE 2025, to appear

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Relative Aetiological Importance of Birth Order and Maternal Age in [redacted]

L. S. Penrose

Proc. R. Soc. Lond. B 1934 **115**, doi: 10.1098/rspb.1934.0051, published 1 August 1934

A METHOD OF SEPARATING THE RELATIVE AETIOLOGICAL EFFECTS OF BIRTH ORDER AND MATERNAL AGE, WITH SPECIAL REFERENCE TO [redacted]

Annals of Eugenics
Vol 6, Issue 1 Oct 1934
pp 108-131

By L. S. PENROSE, M.D.

From the Research Department, Royal Eastern Counties' Institution

"Proc. Roy. Soc. paper has 2 methods of analysis.. "

- Wright's method of partial correlation
  – but it cannot easily handle families of varying sizes, and the mode of their selection.

- These complications are avoided by using the second method, suggested by Prof R. A. Fisher.

  PURPOSE: describe Fisher's method in detail.

Biometrika (2024), **0**, 0, *pp.* 1–8
https://doi.org/10.1093/biomet/asae038
Advance Access publication 8 August 2024

Studies in the history of probability and statistics, LI: the first conditional logistic regression

BY J. A. HANLEY

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 2001 McGill College Avenue, Montréal, Québec H3A 1G1, Canada
james.hanley@mcgill.ca

SUMMARY

Statisticians and epidemiologists generally cite the publications of Prentice & Breslow (1978) and Breslow et al. (1978) as the first description and use of conditional logistic regression, while economists cite the book chapter by Nobel laureate McFadden (McFadden, 1973). We describe the until-now-unrecognized use of, and way of fitting, this model in 1934 by Lionel Penrose and Ronald Fisher.

Some key words: Birth order; Down's syndrome; Estimating equation; Family-based selection; Maternal age; Peer review; Relative odds; Standard error.

e.g. # Applying discrete choice models to predict Academy Award winners

J. R. Statist. Soc. A (2008)
171, *Part* 2, *pp.* 375–394

Iain Pardoe

University of Oregon, Eugene, USA

and Dean K. Simonton

University of California at Davis, USA

His Oscar Predictions http://iainpardoe.com/oscars/

[Received September 2005. Revised June 2007]

Summary. Every year since 1928, the Academy of Motion Picture Arts and Sciences has recognized outstanding achievement in film with their prestigious Academy Award, or Oscar. Before the winners in various categories are announced, there is intense media and public interest in predicting who will come away from the awards ceremony with an Oscar statuette. There are no end of theories about which nominees are most likely to win, yet despite this there continue to be major surprises when the winners are announced. The paper frames the question of predicting the four major awards—picture, director, actor in a leading role and actress in a leading role—as a discrete choice problem. It is then possible to predict the winners in these four categories with a reasonable degree of success. The analysis also reveals which past results might be considered truly surprising—nominees with low estimated probability of winning who have overcome nominees who were strongly favoured to win.

Keywords: Bayesian; Conditional logit; Films; Forecasting; Mixed logit; Motion pictures; Movies; Multinomial logit

Reference to (Nobel Laureate) McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior.

Table 4: Explanatory variables for Best Actress in a Leading Role

1. Indicator for Best Picture Oscar nomination [1939–2004]. Only 25 actresses have won the Best Actress in a Leading Role Oscar for a movie that did not receive a Best Picture nomination (most recently, Charlize Theron for Monster in 2003).

2. Natural logarithm of the number of previous Best Actress in a Leading Role Oscar wins [1938–2004]. 24 percent of Best Actress Oscar nominees with no previous lead actress wins have won the Oscar, whereas 13 percent of Best Actress Oscar nominees with one or more previous lead actress wins have won. This variable has been log-transformed because it is highly skewed.

3. Indicator for winning a Golden Globe for Best Actress in a Leading Role (Drama) [1944–2004]. Of the 62 Best Actress Oscar winners from 1943 to 2004, 31 had won a Golden Globe for Best Actress (Drama) a few weeks earlier.

4. Indicator for winning a Golden Globe for Best Actress in a Leading Role (Musical or Comedy) [1952–2004]. Of the 55 Best Actress Oscar winners from 1950 to 2004, 11 had won a Golden Globe for Best Actress (Musical or Comedy) a few weeks earlier.

5. Indicator for winning a Screen Actor's Guild award [1996–2004]. Of the 11 Best Actress Oscar winners since 1994, eight had already won a SAG award.

Chance 2005, vol 18(4), 32-39

Why not regular (unconditional) logistic regression?

Why not regular (unconditional) logistic regression?

- Data are organized by competition & year (‘set’)

Why not regular (unconditional) logistic regression?

- Data are organized by competition & year ('set')
- There's a winner in each competition [~~indep. Bernoulli r.v.s~~]

Why not regular (unconditional) logistic regression?

- Data are organized by competition & year ('set')
- There's a winner in each competition [~~indep. Bernoulli r.v.s~~]
- Some elements of profile did not exist in earlier years

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

		Profile				Rel. Prob	Prob. Win	Winner?	LogLik
Year	Nominee	X_1	X_2	...	X_K	$e^{\mathbf{X}\beta}$	(P)	(Y)	$(Y \log P)$
2025	Nominee_1	✓	✓	✓	✓	ω_1	$\omega_1 / \sum \omega$	0	-
2025	Nominee_2	✓	✓	✓	✓	ω_2	$\omega_2 / \sum \omega$	0	-
2025	Nominee_3	✓	✓	✓	✓	ω_3	$\omega_3 / \sum \omega$	0	-
2025	Nominee_4	✓	✓	✓	✓	ω_4	$\omega_4 / \sum \omega$	1	$\log P_4$
2025	Nominee_5	✓	✓	✓	✓	ω_5	$\omega_5 / \sum \omega$	0	-

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

		Profile				Rel. Prob	Prob. Win	Winner?	LogLik
Year	Nominee	X_1	X_2	...	X_K	$e^{\mathbf{x}\beta}$	(P)	(Y)	$(Y \log P)$
2025	Nominee_1	✓	✓	✓	✓	ω_1	$\omega_1 / \sum \omega$	0	-
2025	Nominee_2	✓	✓	✓	✓	ω_2	$\omega_2 / \sum \omega$	0	-
2025	Nominee_3	✓	✓	✓	✓	ω_3	$\omega_3 / \sum \omega$	0	-
2025	Nominee_4	✓	✓	✓	✓	ω_4	$\omega_4 / \sum \omega$	1	$\log P_4$
2025	Nominee_5	✓	✓	✓	✓	ω_5	$\omega_5 / \sum \omega$	0	-
						$\sum \omega$	1		
2024	Nominee_1	✓	✓	✓	✓			0	-
2024	Nominee_2	✓	✓	✓	✓	etc	etc	1	$\log P_2$
2024	Nominee_3	✓	✓	✓	✓			0	-

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

		Profile				Rel. Prob	Prob. Win	Winner?	LogLik
Year	Nominee	X_1	X_2	...	X_K	$e^{\mathbf{x}\beta}$	(P)	(Y)	$(Y \log P)$
2025	Nominee_1	✓	✓	✓	✓	ω_1	$\omega_1 / \sum \omega$	0	-
2025	Nominee_2	✓	✓	✓	✓	ω_2	$\omega_2 / \sum \omega$	0	-
2025	Nominee_3	✓	✓	✓	✓	ω_3	$\omega_3 / \sum \omega$	0	-
2025	Nominee_4	✓	✓	✓	✓	ω_4	$\omega_4 / \sum \omega$	1	$\log P_4$
2025	Nominee_5	✓	✓	✓	✓	ω_5	$\omega_5 / \sum \omega$	0	-
						$\sum \omega$	1		
2024	Nominee_1	✓	✓	✓	✓			0	-
2024	Nominee_2	✓	✓	✓	✓	etc	etc	1	$\log P_2$
2024	Nominee_3	✓	✓	✓	✓			0	-
				etc					
1938	Nominee_1	✓	✓		✓			0	-
1938	Nominee_2	✓	✓		✓	etc	etc	0	-
1938	Nominee_3	✓	✓		✓			1	$\log P_3$
1938	Nominee_4	✓	✓		✓			0	-

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

		Profile				Rel. Prob	Prob. Win	Winner?	LogLik
Year	Nominee	X_1	X_2	...	X_K	$e^{\mathbf{x}\beta}$	(P)	(Y)	$(Y \log P)$
2025	Nominee_1	✓	✓	✓	✓	ω_1	$\omega_1 / \sum \omega$	0	-
2025	Nominee_2	✓	✓	✓	✓	ω_2	$\omega_2 / \sum \omega$	0	-
2025	Nominee_3	✓	✓	✓	✓	ω_3	$\omega_3 / \sum \omega$	0	-
2025	Nominee_4	✓	✓	✓	✓	ω_4	$\omega_4 / \sum \omega$	1	$\log P_4$
2025	Nominee_5	✓	✓	✓	✓	ω_5	$\omega_5 / \sum \omega$	0	-
						$\sum \omega$	1		
2024	Nominee_1	✓	✓	✓	✓			0	-
2024	Nominee_2	✓	✓	✓	✓	etc	etc	1	$\log P_2$
2024	Nominee_3	✓	✓	✓	✓			0	-
				etc					
1938	Nominee_1	✓	✓		✓			0	-
1938	Nominee_2	✓	✓		✓	etc	etc	0	-
1938	Nominee_3	✓	✓		✓			1	$\log P_3$
1938	Nominee_4	✓	✓		✓			0	-
	DATA in Black								$\sum LogLik$

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

		Profile				Rel. Prob	Prob. Win	Winner?	LogLik
Year	Nominee	X_1	X_2	...	X_K	$e^{\mathbf{x}\beta}$	(P)	(Y)	$(Y \log P)$
2025	Nominee_1	✓	✓	✓	✓	ω_1	$\omega_1 / \sum \omega$	0	-
2025	Nominee_2	✓	✓	✓	✓	ω_2	$\omega_2 / \sum \omega$	0	-
2025	Nominee_3	✓	✓	✓	✓	ω_3	$\omega_3 / \sum \omega$	0	-
2025	Nominee_4	✓	✓	✓	✓	ω_4	$\omega_4 / \sum \omega$	1	$\log P_4$
2025	Nominee_5	✓	✓	✓	✓	ω_5	$\omega_5 / \sum \omega$	0	-
						$\sum \omega$	1		
2024	Nominee_1	✓	✓	✓	✓			0	-
2024	Nominee_2	✓	✓	✓	✓	etc	etc	1	$\log P_2$
2024	Nominee_3	✓	✓	✓	✓			0	-
				etc					
1938	Nominee_1	✓	✓		✓			0	-
1938	Nominee_2	✓	✓		✓	etc	etc	0	-
1938	Nominee_3	✓	✓		✓			1	$\log P_3$
1938	Nominee_4	✓	✓		✓			0	-
									$\sum LogLik$

DATA in Black

FITTING in Red

$$\Downarrow$$

$$\hat{\beta} = argmax_\beta \sum LogLik$$

Data, (relative & scaled-to-sum-to-1, modelled) Win Probabilities, LogLikelihood Contributions

Year	Nominee	Profile X_1	X_2	...	X_K	Rel. Prob $e^{\mathbf{X}\beta}$	Prob. Win (P)	Winner? (Y)	LogLik ($Y \log P$)
2025	Nominee_1	✓	✓	✓	✓	ω_1	$\omega_1 / \sum \omega$	0	-
2025	Nominee_2	✓	✓	✓	✓	ω_2	$\omega_2 / \sum \omega$	0	-
2025	Nominee_3	✓	✓	✓	✓	ω_3	$\omega_3 / \sum \omega$	0	-
2025	Nominee_4	✓	✓	✓	✓	ω_4	$\omega_4 / \sum \omega$	1	$\log P_4$
2025	Nominee_5	✓	✓	✓	✓	ω_5	$\omega_5 / \sum \omega$	0	-
						$\sum \omega$	1		
2024	Nominee_1	✓	✓	✓	✓			0	-
2024	Nominee_2	✓	✓	✓	✓	etc	etc	1	$\log P_2$
2024	Nominee_3	✓	✓	✓	✓			0	-
				etc					
1938	Nominee_1	✓	✓		✓			0	-
1938	Nominee $_2$	✓	✓		✓	etc	etc	0	-
1938	Nominee_3	✓	✓		✓			1	$\log P_3$
1938	Nominee_4	✓	✓		✓			0	-
	DATA in Black								$\sum LogLik$

FITTING in Red

$\Downarrow$

$\hat{\beta} = argmax_\beta \sum LogLik$

$\Downarrow$

Predictions for 2026: compute $e^{X\hat{\beta}}$ for each 2026 nominee, and rescale to P's

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Relative Aetiological Importance of Birth Order and Maternal Age in [redacted]

L. S. Penrose

Proc. R. Soc. Lond. B 1934 **115**, doi: 10.1098/rspb.1934.0051, published 1 August 1934

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Relative Aetiological Importance of Birth Order and Maternal Age in [redacted]

L. S. Penrose

Proc. R. Soc. Lond. B 1934 **115**, doi: 10.1098/rspb.1934.0051, published 1 August 1934

First submission received by the Royal Society on November 25, 1933.

PROCEEDINGS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Relative Aetiological Importance of Birth Order and Maternal Age in [redacted]

L. S. Penrose

Proc. R. Soc. Lond. B 1934 **115**, doi: 10.1098/rspb.1934.0051, published 1 August 1934

First submission received by the Royal Society on November 25, 1933.

217 families
(210 had 1 affected child, 7 had 2: $\rightarrow$ 224 'Cases')

Peer Review: 1933

Peer Review: 1933

8th December 1933.

Dr. Penrose,
Royal Eastern Counties Institution,
Essex Hall,
Colchester.

Dear Dr. Penrose,

I have had your paper on age and birth order among [redacted] sent to me as referee by the Royal Society, and as I have a good deal to say, I am writing directly to you, instead of letting it trickle through anonymously as extracts from the referee's report. Either way I am afraid you will find it a confounded nuisance.

The whole difficulty turns on the point made in section three, but that section makes it far from clear. You do not mention the essential point, that choosing families only containing [redacted], the proportion of [redacted] must be highest in the smallest families, which generally contain early, but not late children by birth rank.

Now it seems to me that your family data are much too important for you to be satisfied with an unconvincing statistical analysis.

I mean, that no one reading your paper critically will feel sure that a more exact treatment would not have yielded a different result.

I may add that I entirely expect your actual conclusions to be the right ones, but that is no sufficient reason why they should not be adequately established.

or $pq' : p'q$

or $\frac{1}{4}\ \frac{o}{q} : \frac{1}{2}\ \frac{p'}{q'}$

or say, $x : x'$

where x is proportional to p/q .

For families containing one [redacted]

but more than one normal, the expectations are

clearly $\frac{x}{S(x)}$, $\frac{x'}{S(x)}$, $\frac{x''}{S(x)}$

where $S(x)$ is the sum of the values x for the different maternal ages in the family.

For families containing two [redacted] the expectations of [redacted]s at each place will be

$$\frac{x\ S'(x)}{SS(x\ x')}$$

adding up to two.

When $S'(x)$ is the sum of the other values, and $SS(x\ x')$ stands for the sum of all the products two at a time.

Revised Manuscript

"To avoid these sources of ambiguity the data have been subjected to analysis by an entirely different method which was suggested by Professor R. A. Fisher. By use of this new process we are able, after a single complex reconstruction, [JH 2024: conditional logistic model]

Revised Manuscript

"To avoid these sources of ambiguity the data have been subjected to analysis by an entirely different method which was suggested by Professor R. A. Fisher. By use of this new process we are able, after a single complex reconstruction, [JH 2024: conditional logistic model]

to compare the observed number of Down's syndrome cases in any given birth rank with the number which is to be expected on the hypothesis that the probability of a Down's syndrome child depends upon maternal age (in some manner unknown prior to the data) but not, given age, upon birth rank."

Revised Manuscript

"To avoid these sources of ambiguity the data have been subjected to analysis by an entirely different method which was suggested by Professor R. A. Fisher. By use of this new process we are able, after a single complex reconstruction, [JH 2024: conditional logistic model]

to compare the observed number of Down's syndrome cases in any given birth rank with the number which is to be expected on the hypothesis that the probability of a Down's syndrome child depends upon maternal age (in some manner unknown prior to the data) but not, given age, upon birth rank."

i.e., they didn't fit a model with both age and birth order; they fitted one **based just on age**, and then (effectively) **grouped the predictions by birth order**.

Table 2. *Test of theory that birth order is not an aetiological factor: observed (O) and expected (E) numbers of Down's syndrome children, and standard errors*

				Standard Error	
Birth rank	O	E	Difference	1934	2024
1st	26	23.97	+2.03	2.68	2.9
2nd or 3rd	55	57.56	-2.56	3.85	3.9
4th, 5th or 6th	59	61.98	-2.98	4.07	4.1
7th to 10th	61	58.37	+ 2.63	3.41	3.3
11th to 17th	23	22.14	+0.86	1.84	1.8
Total	224	224.02			

Source: Table X page 122 of Penrose (1934a). E's computed from fitted x function in Table 1. 2024 Standard errors estimated by simulation (see text).

Their fitting of their age-only model, where they categorized age as 7 age-bins, each 5 years wide. So their model had 6 free age-effect parameters.

Table 1. *Trial ω values, and age-specific fitted ('calculated') frequencies of Down's syndrome (DS) children*

Data	Maternal age group	15–19	20–24	25–29	30–34	35–39	40–44	45–49
	Observed no. of normal children	10	114	199	228	170	67	15
	Observed no. of DS children	3	13	14	27	64	81	22
Fitting Trial no.								
1	ω values	83[4]	31[2]	19[**1**]	33[2]	104[5]	321[15]	407[20]
	Calculated no. of DS children **(fitted)**	3.69	16.87	19.50	29.11	59.48	76.99	18.35
⋮								
7	ω values	22[4]	10[2]	6[**1**]	19[3]	88[15]	296[50]	558[90]
	Calculated no. of DS children	2.98	12.87	13.82	26.58	64.14	81.46	22.17
.....	.....							
`clogit`	Scaled ω values	[3.47]	[1.59]	[**1**]	[2.96]	[13.19]	[43.62]	[81.53]

Source: Page 440 of Penrose (1934a).

Since the ω values are relative odds, values in brackets have been scaled so that the lowest risk age group (25–29) serves as the reference category, with a scaled odds of **1**:1. See Penrose (1934b) for how he chose the ω's for each trial. The scaled ω values fitted in five iterations by the `clogit` function in the `R survival` package (R Development Core Team, 2024) yielded calculated frequencies that were, in absolute terms, within 10^{-9} of the observed ones.

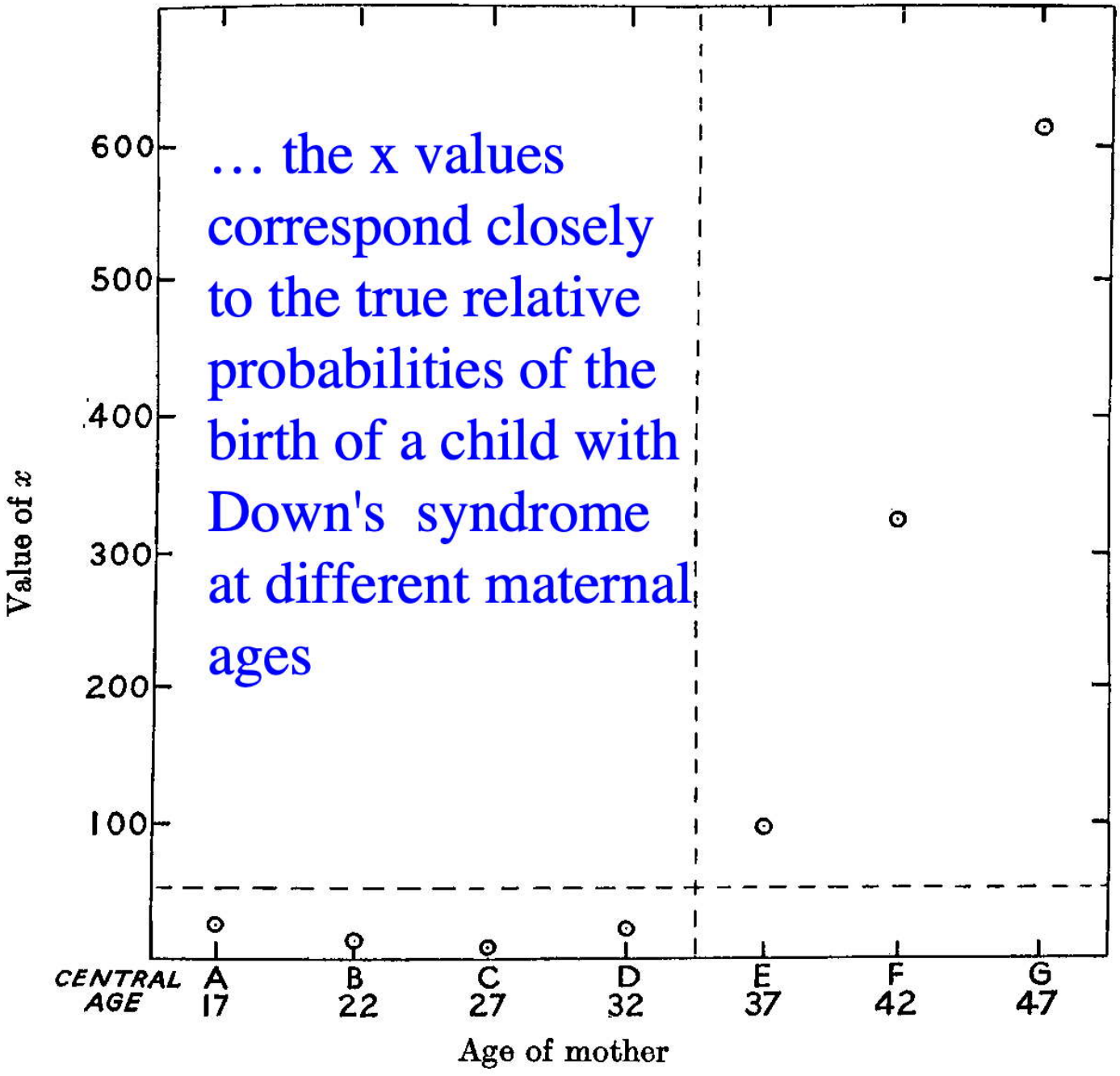


Fig. 1. Final estimate of values of x for maternal age groups

Maternal Age.

Serial number	Sex	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
1	f											4		5		6		7		8		10		11	12		13		**14**				
2	m					**1**		2				3		4																			
3	m			**1**		2			3																								
4	m																					**7**											
5	m					1		**2**																									
6	m, f								2		4			6							8		**9**			**10**							
7	m					1	2		3		4		5			6			7		8		9				**11**						
8	m															1	2		3		**4**		5		6								
9	m																4		5										**7**				
10	m					1							2														**3**						
11	f															**1**	2						3										
12	f							1	2		3			5		6			8			9					12					**13**	
13	f																2		**3**														
14	m														2				**3**														
15	f										2											**3**											
16	f																				1		2		3			**4**					
17	m							1		2		3			4			5		6			**7**							9			
18	f			1			3					4			5	6		7		8				9			**10**						
19	m	1	2	3	4							5									10	11		12			**13**						
20	m, f									1	2	3				4		**5**		6			**7**										
21	f									1		**2**		3													5						
22	m								1					2		3				4		5							**6**				
23	m																			**1**												
24	m																1		2	3		4								**6**			
25	f																								**1**								
26	f							1		2				3					5				**6**										
27	f								1		2		3					5		6												**8**	
28	f												1		2				4						7			**8**					
29	m											1		2		3								**5**						6			
30	m									1						2								3		4			5				

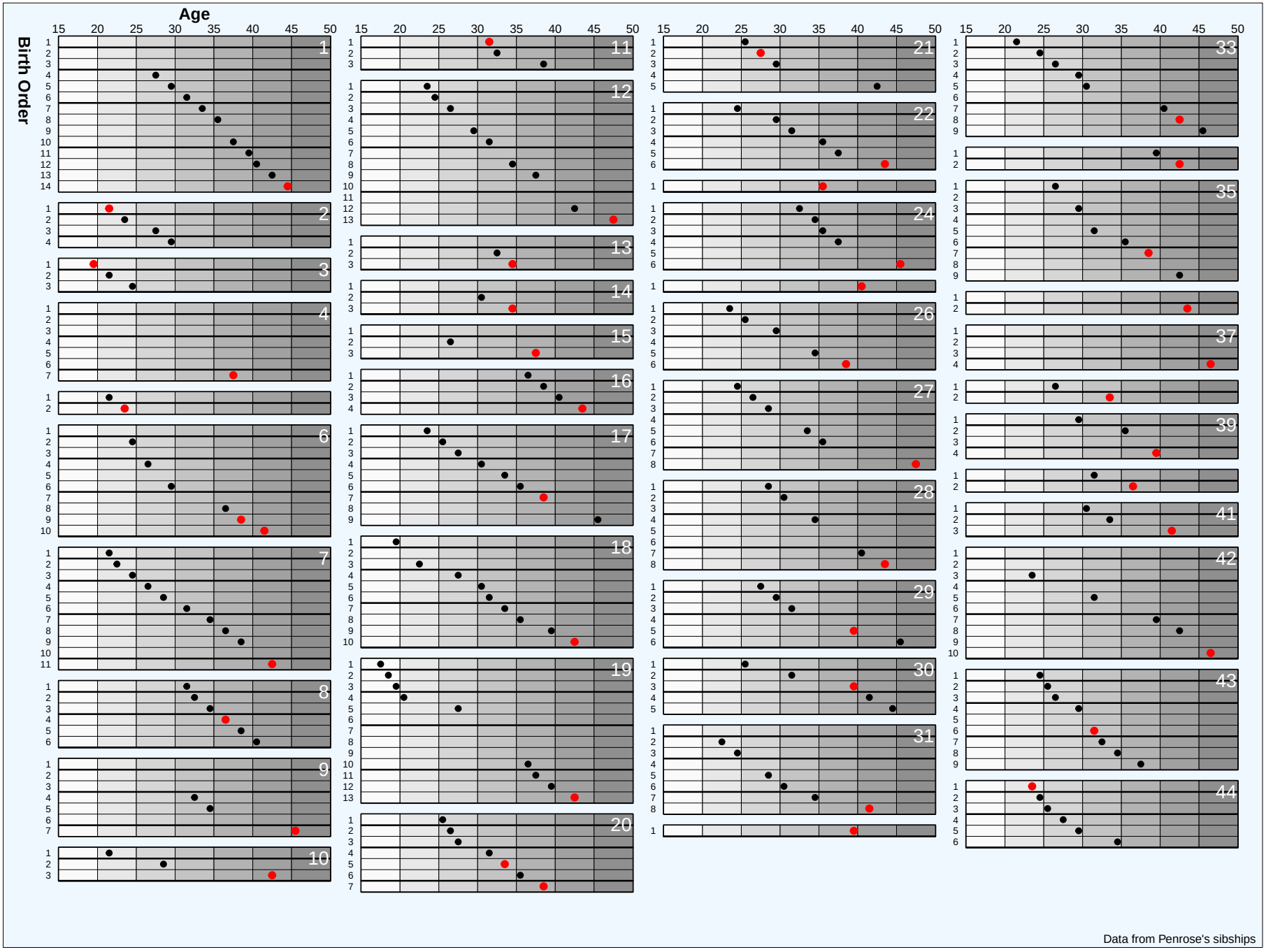
Maternal Age—(*continued*)

Serial number	Sex	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
196	m						1			3			7				**9**																
197	m					1		2	3		4		5		6							**7**										
198	m, f															**3**			4				5						**6**			
199	f								1	2							4	5							**7**								
200	f																					1			2			**3**					
201	m														1		2		3			**4**										
202	f												1						2								**3**						
203	m						**1**																										
204	m													1								**3**											
205	f												1			3	4				5				6							**8**	
206	m			**1**																													
207	f																													**1**			
208	f															1				4		6	7				**8**						
209	f																		1						**2**								
210	m									1		2									3			**4**									
211	m														9	10						13						14				**15**	
212	f												1		**2**	3																	
213	m											4			5		6			**7**													
214	m																		2		3						**6**						
215	m				1		2			3										8		9		10		11		12			13		**14**
216	f															5	6	7			10						14	**15**					
217	m, m																		2											{**3** **3**			
		1	3	6	14	17	16	26	42	39	41	40	38	41	51	42	48	45	42	49	34	28	32	27	21	18	12	11	8	8	6	1	—
		—	—	**3**	**1**	**3**	**4**	**4**	**1**	—	**3**	**6**	**3**	**2**	**2**	**4**	**6**	**6**	**9**	**13**	**8**	**15**	**14**	**14**	**18**	**16**	**23**	**12**	**12**	**7**	**8**	**6**	**1**

Maternal Age—(*continued*)

Serial number	Sex	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
196	m						1			3			7				**9**																
197	m					1		2	3		4		5		6							**7**											
198	m, f															**3**			4				5						**6**				
199	f								1	2							4	5							**7**								
200	f																					1			2			**3**					
201	m														1		2		3			**4**											
202	f												1						2								**3**						
203	m						**1**																										
204	m													1								**3**											
205	f												1			3	4				5				6							**8**	
206	m			**1**																													
207	f																													**1**			
208	f															1				4		6	7				**8**						
209	f																		1						**2**								
210	m									1		2									3			**4**									
211	m														9	10						13						14				**15**	
212	f												1		**2**	3																	
213	m											4			5		6			**7**													
214	m																		2		3						**6**						
215	m				1		2			3										8		9		10		11		12		13			**14**
216	f															5	6	7			10						14	**15**					
217	m, m																		2											{3 3			
		1	3	6	14	17	16	26	42	39	41	40	38	41	51	42	48	45	42	49	34	28	32	27	21	18	12	11	8	8	6	1	—
		—	—	**3**	**1**	**3**	**4**	**4**	**1**	—	**3**	**6**	**3**	**2**	**2**	**4**	**6**	**6**	**9**	**13**	**8**	**15**	**14**	**14**	**18**	**16**	**23**	**12**	**12**	**7**	**8**	**6**	**1**

JH has assembled these data into a ‘long’ .csv file that is available on his website `https://jhanley.biostat.mcgill.ca/Penrose/`


Age
Birth Order
15 20 25 30 35 40 45 50
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Data from Penrose's sibships

CONFIDENTIAL. NOTE.—It is the invariable practice of the Society's Officers, in correspondence with the communicator, and through him with the author of a paper, to preserve the anonymity of a referee. In order, however, to enable an author to consider suggested modifications, or to enable a communicator to consider the desirability of withdrawing a paper, it is frequently desirable to acquaint the communicator with the substance of a referee's criticisms. The referee is requested to *enclose within square brackets* any section of or phrase in his report, which he would wish to be withheld from the communicator in any correspondence of this kind.

REFEREE'S REPORT.

To the Sectional Committee for Physiology.

Paper by L.S. Penrose.

On The relative aetiological importance of birth order and maternal age in Mongolism.

1. Should the paper be read before the Society? ✓

2. Should it be published by the Society? ✓

3. Should it be published in the "Philosophical Transactions" or the "Proceedings"? (*See* Standing Orders on opposite page.) Proceedings

4. Should it be published in full or only in part? Full

5. Are any material modifications necessary? The paper has been reworked so as to meet quite fully the criticism previously ?made

6. Which illustrations, if any, accompanying the paper should be reproduced?

Signature R. A. Fisher

It is requested that any further remarks may be written on foolscap paper.

This paper has been reworked so as to meet quite fully the criticism previously ?made

RECORDS OF SPECIAL CONVOCATIONS DURING THE CONGRESS

10th International Congress of Genetics, McGill University, Aug 20-27, 1958

A SPECIAL CONVOCATION of McGill University was held in the Percival Molson Memorial Stadium on Wednesday, August 20. Degrees of Doctor of Science, *honoris causa*, were conferred upon Professor Hitoshi Kihara, Professor Lionel S. Penrose, and Professor Curt Stern. The remarks of the Principal and Vice-Chancellor, Dr. F. Cyril James, are included below together with the citation of the recipients prepared by Dr. Lloyd G. Stevenson, Dean of the Faculty of Medicine, and the response of Professor Curt Stern for the recipients.

REMARKS OF THE PRINCIPAL AND VICE-CHANCELLOR

It is my privilege this morning, from this Convocation platform, to offer to each of you a warm welcome to Montreal, and especially to McGill University. I hope that the arrangements made for your comfort by the Committee headed by my colleague Professor J. W. Boyes will make your stay pleasant, and that the scientific discussions during this X International Congress of Genetics will be intellectually rewarding. When you meet again, at some other place, for the eleventh Congress, I hope that the tenth will be a pleasant memory to evoke nostalgic talk during the intervals between more serious discussion.

Genetics is a comparatively new science, but we at McGill were early indoctrinated by the enthusiasm and skill of a master. Although Leonard Huskins—whom the Genetics Society of Canada commemorates in its annual Memorial Lecture—came to McGill as Associate Professor of Botany in 1930, his enthusiasm

Special Convocation, McGill University: *left to right*—Professor Sewall Wright, Professor L. S. Penrose, Professor Curt Stern, Professor H. Kihara, Professor F. Cyril James (Principal and Vice-Chancellor)

New York Times, Feb. 9, 1964

Kennedy Foundation

Its Efforts Set an Example for U. S. In Fight Against Mental Retardation

Last Wednesday evening a distinguished group of citizens headed by President Johnson met in New York to pay tribute to three physicians and three civic leaders for their contributions to the fight against mental retardation.

The occasion was the presentation of the Second Annual International Awards of the Joseph P. Kennedy Jr. Foundation.

Those honored were:

Dr. Grover Francis Powers, professor emeritus of pediatrics, Yale University.

Dr. Robert P. L. Lafon, professor of neuropsychiatry, University of Montpellier, France.

Dr. Lionell S. Penrose, professor of eugenics, University College, London.

Senator Lister Hill of Alabama.

Bert T. Combs, former Governor of Kentucky.

Representative John E. Fogarty of Rhode Island.

In many ways the dinner, which was preceded by a day-long scientific program, was a tribute to President Kennedy.

It was during Mr. Kennedy's Presidency that the spotlight was focused for the first time on mental retardation nationally through the President's Panel on Mental Retardation.

$631 Million Provided

As a result of two major pieces of legislation passed last

Joseph P. Kennedy, Jr. Foundation Annual International Awards Dinner New York, N.Y. Program

Sargent Shriver
Executive Director, Joseph P. Kennedy, Jr. Foundation

The Honorable Edward M. Kennedy
President, Joseph P. Kennedy, Jr. Foundation

The Honorable Lyndon B. Johnson
President of the United States

The Right Honourable Lester Bowles Pearson
Prime Minister of Canada

TO SOLVE A HUMAN PUZZLE
Film of the achievements of the International Award Winners
Produced by the Joseph P. Kennedy, Jr. Foundation
in cooperation with the Warner-Lambert Pharmaceutical Co.

Guest Master of Ceremonies: Jack Benny

Special Guest Stars: Ethel Merman, Nat "King" Cole

Music by Count Basie

The problem of mental retardation was once obscured by ignorance and superstition. Today it is brightened by a new spirit of hope. Much of this new hope stems from the studies and writings of *Dr. Lionel S. Penrose,* Galton Professor of Eugenics at the University of London. For more than thirty years, Dr. Penrose has been investigating the genetic factors in mental retardation. His "Colchester Survey" was the first large-scale systematic attempt to identify specific causal factors in a large group of retardates. All of his many books and papers are a provocative blend of terse fact and stimulating theoretical speculation. His work has been—and will continue to be—an inspiration to other dedicated men to attack the problem of mental retardation until, hopefully, it is solved.

LYNDON B. JOHNSON

36th President of the United States: 1963 - 1969

Remarks in New York City at the Joseph P. Kennedy, Jr., foundation Awards Dinner

February 05, 1964

Mr..Chairman, Mr. Prime Minister, Your Eminence, Mrs. Kennedy, members of the Kennedy family, award winners, distinguished guests and friends:

Earlier generations of Americans were fortunate to have known the Adamses, the Lees, the Randolphs, the Lafollettes, and the Roosevelts.

Our generation is proud and blessed to have known the Kennedys. They are an extraordinary family. fierce competitors in life, they are a closely knit team, for they are all genuine friends as well as relatives-united in prayer, devoted to their parents and maintaining a community of purpose and practice which makes them the world's second most powerful Common Market.

Unlike many who have their opportunities, they prefer labor to leisure. They place the public good ahead of private gain. They both preach excellence and pursue it. They have been granted more than their share of greatness, but they have also been dealt more than their share of grief. The senseless, mindless murder of their martyred brother and son brought endless, timeless grief to every American home.

John Kennedy was to have been here tonight. No cause was closer to his heart. Millions of people, at home and around the world, will reap the harvest of his pioneering work in this field--a field which has been so greatly misunderstood and so greatly neglected so many years.

Humbly, I shall carry on for him here, as I intend to carry on the great efforts that he started for lasting peace. With his memory and his spirit to inspire us, and his words and his works to guide us, we shall live up to that trust. We shall finish his fight; and we shall conquer mental retardation and mental illness and poverty and every other foe of the land that he loved, and every other foe of the people he served.

NEW YORK. PRESIDENT JOHNSON PRESENTS ONE OF SIX INTERNATIONAL AWARDS OF THE JOSEPH KENNEDY JR. FOUNDATION TO DR. LIONEL PENROSE (*LEFT*), PROFESSOR OF EUGENICS AT UNIVERSITY COLLEGE, LONDON, FOR HIS CONTRIBUTION IN THE FIGHT AGAINST MENTAL ILLNESS. THIS FOUNDATION WAS ESTABLISHED IN 1946 IN MEMORY OF THE LATE PRESIDENT'S BROTHER, KILLED DURING THE WAR. (*Wire photograph.*)

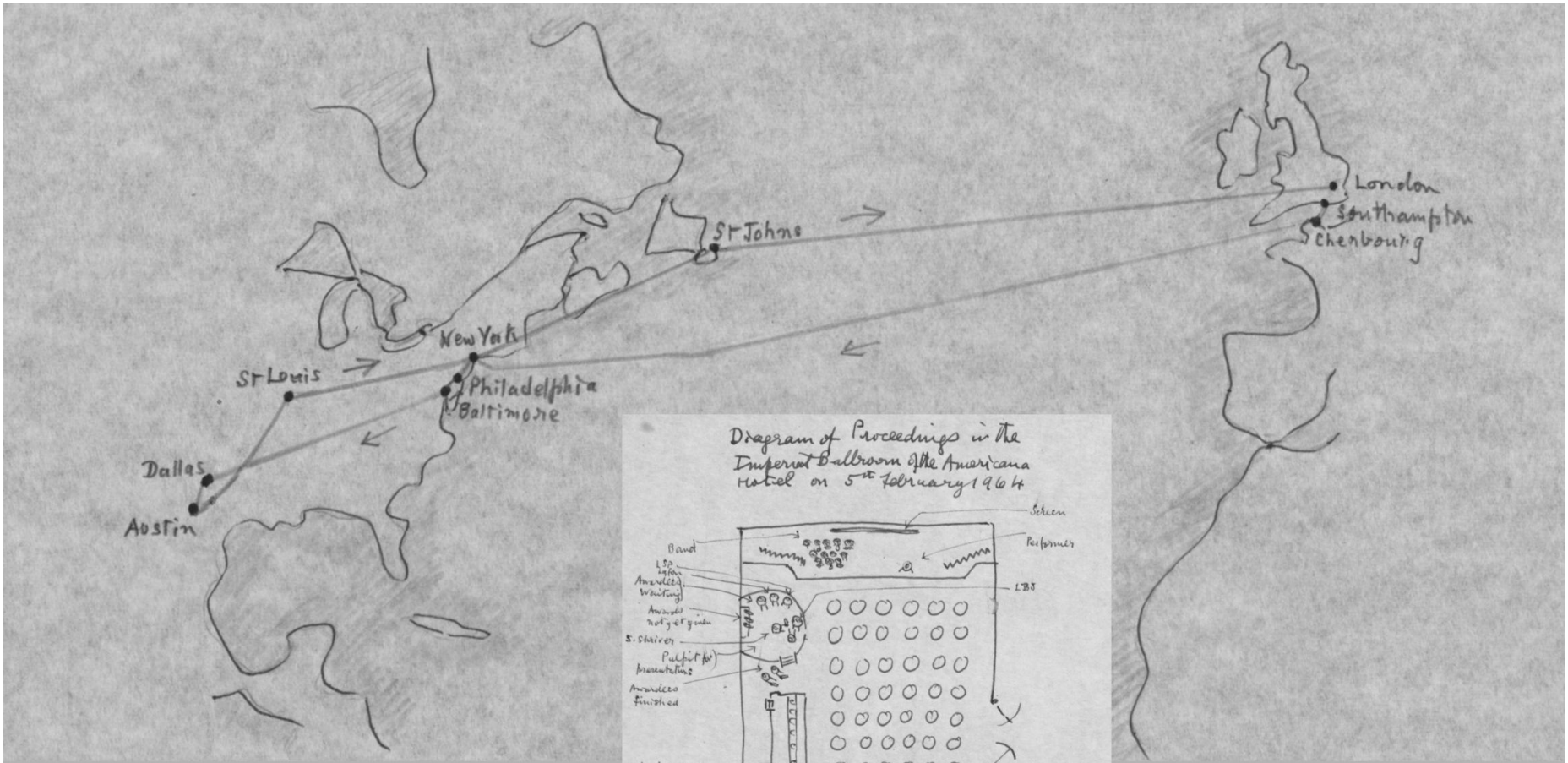
FEBRUARY 15, 1964
THE ILLUSTRATED LONDON NEWS
A WINDOW ON THE WORLD.

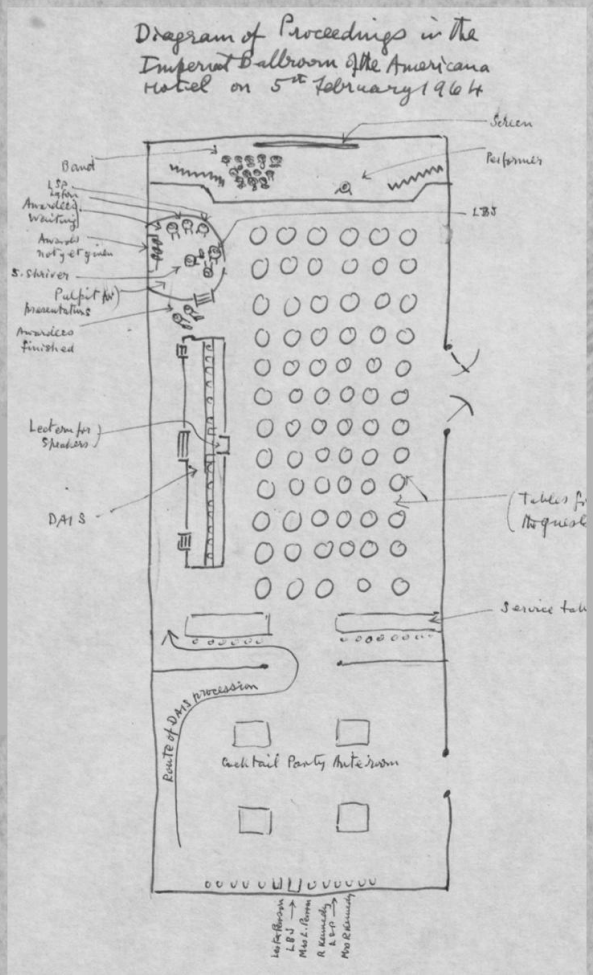
"We can say with some assurance that, although children may be the victims of fate, they will not be the victims of our neglect."

Statement by John F. Kennedy, October 24, 1963, upon signing the Child Health and Mental Retardation Planning Amendment.

Three Trains, Three Planes, a Ship and an Etched Piece of Glass with a Silver Stand

1964





PAPERS

18 pp.

Three Trains, Three Planes, a Ship
and an Etched Piece of Glass with a Silver Stand

1964

29th January to 2nd February

First of all, about the ship : in my opinion the Queen Elizabeth is a quite remarkable object. Since she is 1031 feet

December 6, 1962 - President John F. Kennedy, Awards Dinner of the Joseph P. Kennedy Jr. Foundation

LINK

Material related to Biometrika 2024 article on the conditional logistic regression fitted by Lionel Penrose and Ronald Fisher in 1934.

1934 Articles

* The relative aetiological importance of birth order and maternal age in mongolism.
  Penrose LS. Proceedings of the Royal Society B 115 431-450.
* A method of separating the relative aetological effects of birth order and maternal age, with specific reference to mongolian imbecility.
  Penrose LS. Annals of Eugenics 6 108-132.

Data set

* Penrose's 1934 Data: 217 sibships with at least one child with DS (.csv)
  The dataset was reconstructed from the appendices to the two 1934 articles.
  It has information on 224 affected children (the no. reported in the articles),
  but on 80**6** unaffected children (rather than the 80**7** reported)

Archives

* Birth Order and Down's syndrome Correspondence at Wellcome Library
* Penrose Papers at UCL Digital Collections
* R A Fisher Correspondence Files at University of Adelaide
* Penrose's diary account of the family tour of North America in 1958
  before the 10th International Congress of Genetics at McGill University, and the conferring of an honourary doctorate (last few slides)
* Penrose's diary account of his 1964 visit to the USA, where (in NYC) he received the Joseph P. Kennedy Jr. Foundation's International Award from LBJ
* First Awards ceremony (1962), where the awards were presented by JFK.

Material related to

"The first conditional logistic regression and the first nested case-contol study"

McGill Biostatistics Seminar Series, 2024.11.27

Presentation

video | 'slides' lyrics (these include some material not in video/audio)

Background:

* Reviews: Cole-1979 Liddell-1988 Breslow-1996 Doll-2001 Breslow-2007
* Articles:

Webinar presentation

"A not-well-known Lionel Penrose - Ronald Fisher collaboration"

Sponsored by CHANCE magazine and the ASA History of Statistics Interest Group

September 24, 2025

video, to be added | 'slides' | lyrics

WRAP-UP

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**:

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.
- But need higher-dimensional data to check model fits!

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.
- But need higher-dimensional data to check model fits!

- **1934 paper**:

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.
- But need higher-dimensional data to check model fits!

- **1934 paper**: re-fitting 'no-name' model from 1st principles gives many insights.

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.
- But need higher-dimensional data to check model fits!

- **1934 paper**: re-fitting 'no-name' model from 1st principles gives many insights.
- **2 types of statistical models**:
  - *mathematical*, to be fitted and evaluated (and thought-hard-about)

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.
- But need higher-dimensional data to check model fits!

- **1934 paper**: re-fitting 'no-name' model from 1st principles gives many insights.
- **2 types of statistical models**:
  - *mathematical*, to be fitted and evaluated (and thought-hard-about)
  - *human*, to be studied and emulated (the book title is taken! but . . .)

Several messages

- Despite its imperfections, **broad findings** of 1933 & 1934 studies have **held up**.
- Nowadays, mother's-age 'curve' = **mix** of flat line (indep. of age) + a curve.

- **1933 paper**: sufficient statistics = less (human) computational labour.
- sufficient statistics [+ exponential families] = more data privacy.
- But need higher-dimensional data to check model fits!

- **1934 paper**: re-fitting 'no-name' model from 1st principles gives many insights.
- **2 types of statistical models**:
  - *mathematical*, to be fitted and evaluated (and thought-hard-about)
  - *human*, to be studied and emulated (the book title is taken! but . . .)

- Statistical '**Archaeology**' can be fun – and instructive and inspiring!

Thanks to:

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]
- Penrose (UCL) & Fisher (U Adelaide) Online Archives

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]
- Penrose (UCL) & Fisher (U Adelaide) Online Archives
- Supratik; Serendipity; `lpSolve` (Sam Buttrey)

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]
- Penrose (UCL) & Fisher (U Adelaide) Online Archives
- Supratik; Serendipity; `lpSolve` (Sam Buttrey)
- Andrea Benedetti and her `bios624` class, Fall 2014

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]
- Penrose (UCL) & Fisher (U Adelaide) Online Archives
- Supratik; Serendipity; `lpSolve` (Sam Buttrey)
- Andrea Benedetti and her `bios624` class, Fall 2014
- All of my colleagues since 1973

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]
- Penrose (UCL) & Fisher (U Adelaide) Online Archives
- Supratik; Serendipity; `lpSolve` (Sam Buttrey)
- Andrea Benedetti and her `bios624` class, Fall 2014
- All of my colleagues since 1973
- My (trained-at-Rothamsted under Frank Yates) University College Cork statistics professor Tadgh Carey who said to our small (1966-1969) class

Thanks to:

- ChatGPT [Referee's Report on the Revised Ms.]
- Penrose (UCL) & Fisher (U Adelaide) Online Archives
- Supratik; Serendipity; `lpSolve` (Sam Buttrey)
- Andrea Benedetti and her `bios624` class, Fall 2014
- All of my colleagues since 1973
- My (trained-at-Rothamsted under Frank Yates) University College Cork statistics professor Tadgh Carey who said to our small (1966-1969) class

  "You are still too young but 'one day' I will let you see the statistics journals where Fisher and Pearson were so nasty to each other."