Hilda Mary Woods and the negative binomial distribution

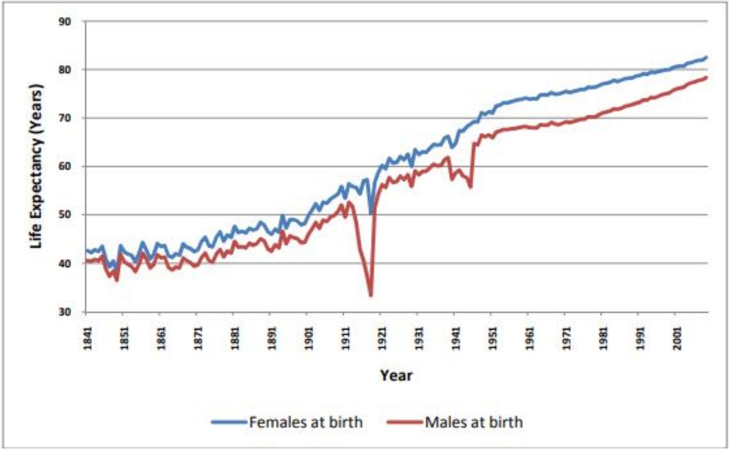
Mario Cortina Borja

Population, Policy and Practice Research and Teaching Department
Great Ormond Street Institute of Child Health
University College London

8th March 2021

Overview

A decade of wars: life expectancy in England & Wales, 1841–2009



Source: England and Wales, Total Population, Life tables (period 1x1), Males and Females. Last modified: 03-Nov-2010, MPv5 (May07).
© Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany).
(www.mortality.org)

Source: Pattison J et al (2013)

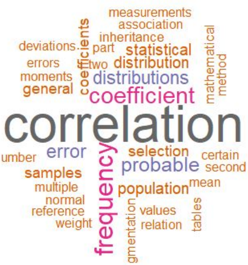
Biometrika and *JRSS*: 1910–1920

- *Biometrika*
  - Founded in 1901 by K Pearson, R Weldon, and F Galton
  - Pearson named it and FY Edgeworth suggested the ‘k’ in the name
  - See Cox (2001), and Aldrich (2013) for a history of the early years
- *JRSS*
  - Founded in 1838 as *Journal of the Statistical Society of London*
  - Renamed *Journal of the Royal Statistical Society* in 1887
  - Divided into *Series A (General)* and *Series B (Statistical Methodology)* in 1948)
  - *Series C (Applied Statistics)* started in 1952
  - *Series A* renamed *Series A (Statistics in Society)* in 1988
  - *Series D (The Statistician)* 1993 after merge of the Society with the Institute of Statisticians
  - *Series D* replaced by *Significance* from 2004
  - See Hill (1984) for a history of the early years

Biometrika and *JRSS*: 1910–1920

	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	Total
BKA	0	38	0	24	37	24	0	0	19	0	21	163
JRSS	28	27	22	31	21	22	19	27	14	11	20	242

Biometrika: 1910-1920



JRSS: 1910-1920

Biometrika and *JRSS*: 1910–1920

Author	JRSS	BKA	Notes
[Anonymous]	41	14	Several signed 'K.P.'
Pearson, K (1857–1936)	2	52	
Edgeworth, FY (1845–1926)	11	0	
Elderton, EM (1878–1954)	1	10	collaborated with WP Elderton
Snow, EC (1886–1959)	7	3	
Greenwood, M (1880–1949)	6	2	
Isserlis, L (1881–1966)	1	6	
Bowley, AL (1869–1957)	6	0	
Harris, JA (1880–1936)	0	6	
Lehfeldt, RA (1868–1927)	6	0	
Paish, G (1867–1957)	6	0	
Welton, TA (1835–1918)	6	0	
Yule, GU (1871–1951)	6	0	
Baines, JA (1847–1925)	5	0	
Lee, A (1858–1939)	1	4	Evolutionary biologist; worked in munitions Department in WW1
Bell, J (1879–1979)	0	4	Geneticist
Wood, F (1883–1919)	3	0	First woman in RSS Council
Davin, AG	0	2	Assistant in KP's laboratory
Thomson, EY	0	2	Assistant in KP's laboratory
Ryley, KT	0	1	Assistant in KP's laboratory

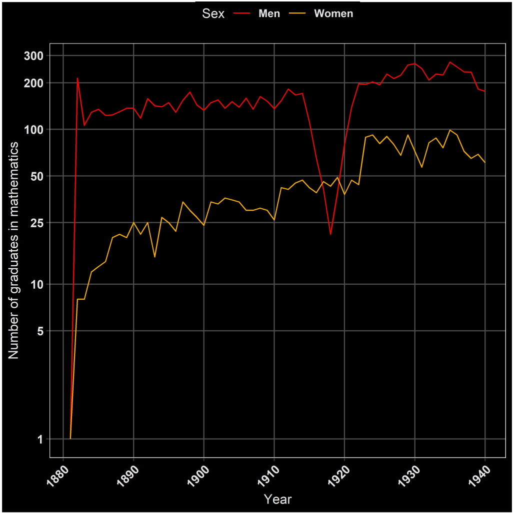
The emergence of modern statistics

- Statistics became a rigorous mathematical subject from the second half of the 19th century: Quetelet, Lexis, Gini, Mayo–Smith...
- In Britain, the 1910s saw developments in medical statistics, experimental design, hypothesis testing and methods for use with small data samples
- In official statistics, for the first time individuals in a household could also make separate confidential returns in the census of 1921 – which included greater detail than any previous census

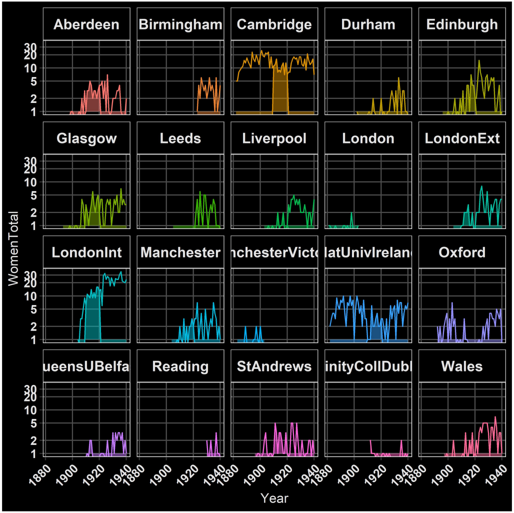
Women and mathematics degrees: the Davis historical archive

- Data collected by AEL Davis, and kept at the History of Mathematics website, St Andrews University
- The archive contains details of the approximately 2500 women who graduated in mathematics from universities in Britain and Ireland before 1940
- See Davis (2017) for a description
- Data available at `https://mathshistory.st-andrews.ac.uk/Davis/`

Women and mathematics degrees: the Davis historical archive

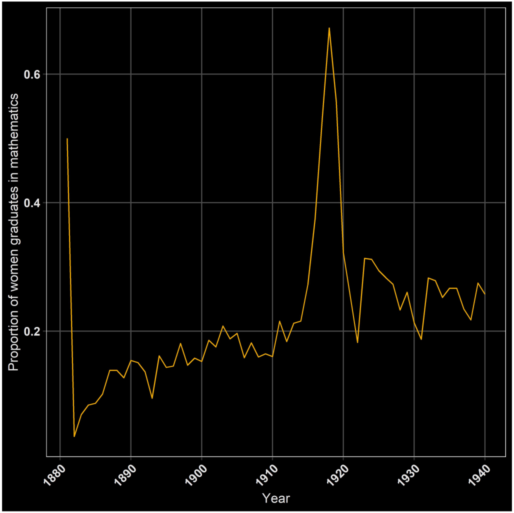


Women and mathematics degrees: the Davis historical archive



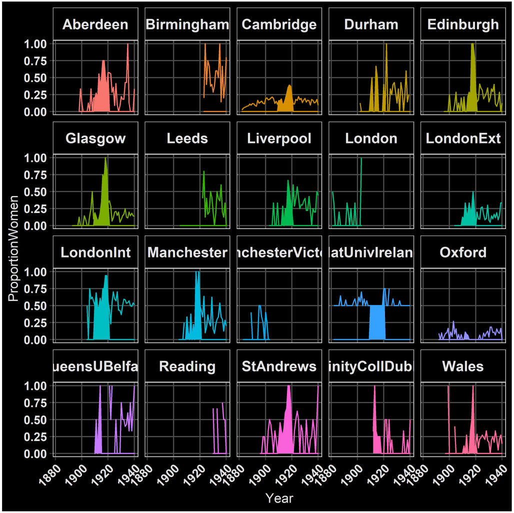
Women and mathematics degrees: the Davis historical archive

UCL

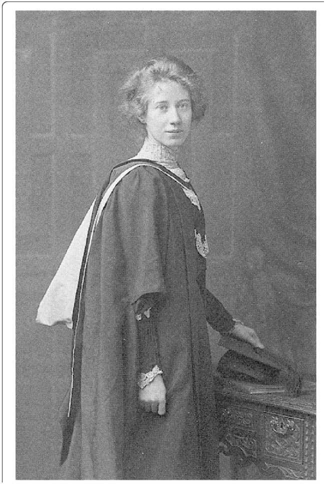


GREAT ORMOND STREET HOSPITAL CHARITY

Women and mathematics degrees: the Davis historical archive



Julia Bell (1879–1979)



source: Bundey S (1996) *J Med Biography*

Julia Bell (1879–1979)

- Took the Mathematical Tripos examination at Cambridge in 1901
- Worked at Cambridge Observatory investigating solar parallax, 1902–1908. Awarded a Master's degree (MA) from Trinity College, Dublin, on this work
- Statistical assistant to Karl Pearson at UCL, 1908–1914
- Diploma (MRCS, LRCP) in medicine from the London School of Medicine for Women, at the Royal Free Hospital, 1914–1920
- Research fellow at UCL Galton laboratory (1920–1933), National Neurological Hospital (1936–1944)
- Retired aged 65 in 1944 but continued to work as an Honorary Research Fellow until 1965
- Many important contributions to medical genetics
- See Bundey (1996) *J Med Biography*

Frances Wood (1883–1919)

source: Cole TJ (2017), *Significance*

Frances Wood (1883–1919)

- Graduated in Chemistry from UCL in 1908 – four sisters were graduates
- Recruited by Major Greenwood in 1912 to work in the newly formed Statistics Department at the Lister Institute
- Seconded to the Board of Trade in 1914; transferred to the Ministry of Munitions in 1916, where she remained until her resignation, in 1919, on account of her pregnancy
- First female member of the RSS Council (1915) and of the RSS Executive Committee (1917); appointed OBE in 1918
- Daughter Barbara (1919–2009) born in 1919
- Died in 1919, aged 35, of septicaemia following caesarean section
- RSS established The Frances Wood Memorial Prize in 1920; restored in 2017, as The Wood Medal
- See Cole TJ (2017) *Significance*

Frances Wood (1883–1919)

Proposed Memorial

TO

Mrs. FRANCES WOOD

WHEN the untimely death of Mrs. Frances Wood (née Chick), on October 12th, became known, many of those who had had the privilege of working with her at University College, at the Lister Institute, at the Statistical Society or in Government service at the Board of Trade or the Ministry of Munitions, felt a desire to express their appreciation of her charming personality and her remarkable intellectual ability in some form of permanent memorial. It has been suggested that the most suitable form for such a tribute would be the endowment of, say, a biennial prize–entitled the Frances Wood Memorial Prize–which should be worth, if possible, £20, to be awarded to the author of the best investigation of a sociological problem on statistical lines produced within the two years previous to the award.

If you approve of this scheme and desire to join with us in it, any of the undersigned would be glad to receive contributions.

If there is a sufficient response to this proposal–we suggest a minimum of £200–we will approach the Council of the Statistical Society, who will be asked to undertake the trust and to award the prize.

We enclose a short memoir in appreciation of Mrs. Wood.

(Signed) A. W. FLUX, Royal Statistical Society, 9 Adelphi Terrace, W.C. 2

M. GREENWOOD, Lister Institute, Chelsea Gardens, S.W. 1

W. T. LAYTON, Temple Bar House, 23 Fleet Street, E.C. 4

WINIFRED SMITH, University College, Gower Street, W.C. 1

December, 1919

The Wood Medal

The Wood Medal was established in memory of the statistician Frances Wood OBE (1883–1919). It is awarded every three years to a fellow of the Society for excellent contributions to economic or social statistics.

2021 recipient: Katie Harron
The Wood Medal is awarded to Katie Harron (University College London) in recognition of her outstanding methodological work on record linkage. Read the full citation.

2017 recipient: George Leckie
George Leckie was awarded the Wood Medal in 2017 for his work analysing school performance data and development of procedures for modelling variance structures. His substantive results include ways of visualising the inherent uncertainty surrounding school rankings.

source: Cole TJ (2017), *Significance*; and https://rss.org.uk/training-events/events/honours/wood-medal/

Hilda Mary Woods

- Educated at home and at Girls' s High School, Nottingham
- Litentiate of Royal Academy of Music, in piano, 1909
- Fiancée killed in WW1, 1915
- Meets Major Greenwood in 1916
- Investigates Tuberculosis and industrial conditions, 1916–1919; this leads to the report on accidents co–authored with Greenwood

Hilda Mary Woods

MEDICAL RESEARCH COMMITTEE

INDUSTRIAL FATIGUE RESEARCH BOARD

(Later the Industrial Health Research Board)

REPORT No. 4

THE INCIDENCE OF INDUSTRIAL ACCIDENTS UPON INDIVIDUALS

with Special Reference to Multiple Accidents

By MAJOR GREENWOOD and HILDA M. WOODS

We have:—

Accidents per Person.	*Frequency by A.*	*Frequency by B*
0	9608	9606
1	384	388
2	8	6
3	0	0

The reason of this accord is that in each case the form of distribution approximates to that obtained by the series mentioned in the first section of the appendix, viz. :—

$$e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}\dots\right)$$

where λ is the number of accidents divided by the number of persons. Consequently whenever the conditions are such that the chance of sustaining an accident must be deemed small (and these conditions are fulfilled in the majority of cases), the unmodified pigeon hole schema is probably a quite adequate test of the likelihood that the distribution is one of pure chance in origin, and this is the test we have uniformly applied.

3.—*Distribution of Unequal Liabilities* (U.D.).

Supposing the distribution of susceptibility to accidents to be continuous and of the form :—

$$y = y_0\, e^{-c\lambda} .\ \lambda^{r-1} \quad \dots \quad \dots \quad \dots \quad (5)$$

where λ is the measure of liability or susceptibility c, r and y_0 constants ;

Then the frequencies of 0, 1, 2, &c. accidents are given by the successive terms of :—

$$N\left(\frac{c}{c+1}\right)^r\left\{1+\frac{r}{c+1}+\frac{r(r+1)}{2!(c+1)^2}+\frac{r(r+1)(r+2)}{3!\ (c+1)^3}+\dots\right\} \quad (6)$$

and r and c are obtained from the statistics from :—

$$M = \frac{r}{c} \quad \dots \quad \dots \quad \dots \quad (7)$$

$$\mu_2 = \frac{r(c+1)}{c^2} \quad \dots \quad \dots \quad \dots \quad (8)$$

Where M is the mean and μ_2 the second moment about the mean of the distribution observed.

Hilda Mary Woods

- Moves, following Greenwood, from the Lister Institute to Ministry of Health's Statistical Services, 1919
- Co–authors paper in *Biometrika* on the height and weight of patients in mental hospitals, 1925
- Fellow of the RSS, 1926; remains a Fellow of the Society until 1951
- Joins the Division of Epidemiology & Vital Statistics at the London School of Hygiene and Tropical Medicine (LSHTM) as assistant lecturer in 1928 (the first female lecturer at LSHTM)

Hilda Mary Woods

- Permanent Lecturer at LSHTM, 1928–1933: writes papers on respiratory diseases, scarlet fever, and diphteria, and on methodological paper on interpolation in life tables
- Publishes, with William Russell, *An Introduction to Medical Statistics*, one of the first textbooks on the subject, 1931
- Awarded a DSc by the University of London, 1933
- Her engagement, in 1933, means end of tenure at LSHTM
- Moves to Ceylon to marry Roger Fowke in December 1933; her

Hilda Mary Woods

From Farewell, Johnson and Gear (2012):

- In her late 30s Hilda Woods wrote a short document concerning her start in medical statistics in 1916, titled *The birth of a medical statistician*
- It starts thus: “I am nearly forty, supposed to be quite good looking and my dressmaker seems to enjoy making me clothes because “madam carries them well”. I began work on statistics when I was a flapper.”
- Hilda’s description of herself as a “flapper” was as one who believed in equality between the sexes, and dancing the foxtrot. Greenwood uses the word in a descriptive, and most definitely not in a derogatory sense.

Hilda Mary Woods

- Works at the Medical Department of the Ceylon Government studying prevalence of diseases, and lectures at University College, Colombo, 1934–1938
- Appointed MBE, 1936
- Returns to England, 1938; joins a research team at the the Ministry of Food to study vitamins and minerals
- Member of the Nutrition Society, 1945
- Emigrates to Kenya, 1946
- Adopts her niece Rosemary, whose mother died the day she was born, 1947
- Lives in New Zealand, Australia, Rhodesia, and returns to England in 1958
- Moves to South Africa, 1970, and dies there in 1971
- See Farewell, Thompson & Gear (2012) *JRSS–A*

Hilda Mary Woods

source: Farewell & Johnson (2010) *Statistics in Medicine*, by pemission of Mrs Rosemary Gear.
Portrait of Hilda Mary Woods in her University of London Doctorate of Science robes, May 1933.

Greenwood and Yule's paper (1920)

Major Greenwood and G. Udny Yule (1920)
An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society*, **83**: 255–279.

- Constructs statistical models to fit empirical distributions of accidents in munitions factories during WWI
- Data collected and firstly analysed by Miss Hilda Woods

Greenwood and Yule's paper (1920)

- Though difficult to follow, Greenwood and Yule, derive the negative binomial distribution from the Poisson model
- A clearer, more succinct derivation appears in the 1919 Report by Greenwood and Woods

(1) 648 *Women working on 6-inch H.E. shells for five weeks.*

No. of accidents.	Observations.	Simple Poisson series.	Method of equation (8).	Method of equations (51) and (52).
0	447	406	452	442
1	132	189	117	140
2	42	45	56	45
3	21	7	18	14
4	3	1	4	5
5	2	0·1	1	2

Overdispersion

- Overdispersion occurs when the mean is larger than the variance: this means that as well as a parameter to model location, we need at least one extra parameter to model relatively long tails
- For overdispersed data, it is an excess of individual variability with respect to a Poisson model

Parameter–mix models

- A parameter–mix distribution for a r.v. Y is defined by a pdf dependent on the parameters $\theta_1\,\theta_2,\cdots,\theta_m$ where some (or all) of those parameters are random variables varying according to other distribution
- Then, $f_Y\left(y;\theta_1,\theta_2\cdots,\theta_m\right)$
- If only one parameter θ varies: $Y \sim f_A \bigwedge_\Theta f_B$
- f_A = original distribution, f_B = distribution of the r.v for θ, i.e. the mixing distribution with parameter space Θ

Parameter mix: negative binomial

Negative Binomial = Poisson $\bigwedge_{\Lambda}$ Gamma

- $Y \sim$ Poisson (λ) means that a each individual in a sample of size n, $\{Y_1, \cdots, Y_n\}$ takes values from that distribution
- Thus they have a common mean (λ), and variance (also λ in the Poisson model)

Negative Binomial = Poisson $\bigwedge_{\Lambda}$ Gamma

Often we can assume that the parameter from the Poisson from which each individual's value is drawn is itself a r.v. **corresponding to unobserved frailty, or propensity**

Parameter mix: negative binomial

Negative Binomial = Poisson $\bigwedge_\Lambda$ Gamma

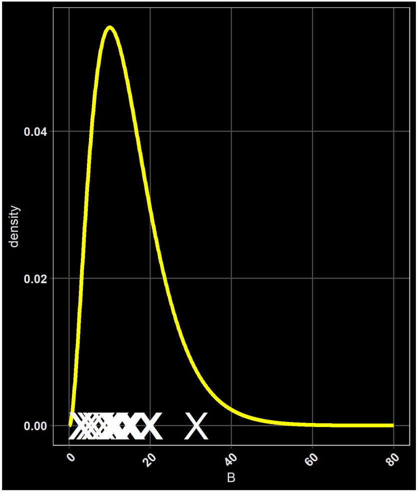
- This is similar to a random effect term in mixed–effects models with a Normal outcome
- However the distribution of the parameter λ, (i.e. f_Λ) must take non–negative values only
- The negative binomial distribution results from mixing a Gamma density with a Poisson r.v.

Parameter mix: negative binomial

Negative Binomial = Poisson $\bigwedge_{\Lambda}$ Gamma

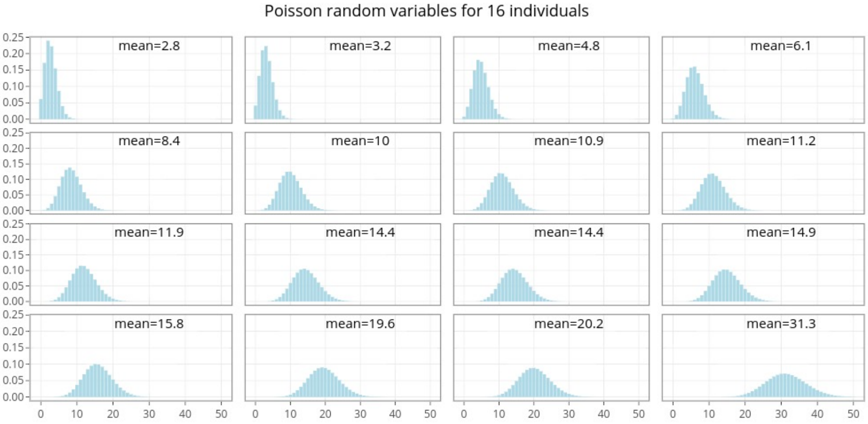
- Greenwood and Yule (1920) construct (without naming it thus) the negative binomial using such a mixing argument (without naming it thus)
- There are many other ways of constructing the negative binomial, notably as the outcome of an urn problem: it measures the number of draws before a numbe of failures Y occurs
- It's a hugely versatile distribution

Parameter mix: negative binomial



A sample of 16 values for $\Lambda \sim \Gamma(3, 1/5)$

Parameter mix: negative binomial



Poisson distributions for 16 individuals with frailty $\sim \Gamma(3, 1/5)$

Parameter mix: other distributions

- Recall that Negative Binomial = Poisson $\bigwedge_{\Lambda}$ Gamma
- Algebraically, parameter–mix distributions are obtained by integrating out Λ from the random effects of the mixing distribution
- Many models have been constructed based on this principle – though there are other forms to obtain the same resuting distributions, e.g. urn problems, or as sums of correlated Poisson r.v.s
- Two well–known examples assume the parameter in the Poisson model to follow an Inverse Gaussian (IG) or a Generalized Inverse Gaussian distribution (GIG), which can have larger tails than the Γ – the resulting models are the **Hermite** and **Sichel** distributions
- Regression Hermite and Sichel models (and value–inflated Sichel) can be fitted with libraries `hermite` and `gamlss` in R

Example: Accidents in machinists in a six–month study (Greenwood & Woods (1919)

value	observed	Poisson	NegBin	Sichel
0	69.00	52.99	66.77	74.39
1	54.00	69.85	60.56	63.49
2	43.00	46.04	36.91	34.27
3	15.00	20.23	18.83	15.39
4	13.00	6.67	8.67	6.35
5	1.00	1.76	3.74	2.52
6	2.00	0.39	1.54	0.98
7	1.00	0.07	0.61	0.38
BIC		632.02	623.85	629.71
X^2		34.90	7.10	14.00
p–value		0.00	0.07	0.01

Data collected by Hilda Woods (MRC report, 1919)

- Aldrich J (2013) Karl Pearson's *Biometrika*, 1901–1936. *Biometrika*, **100**, 3–15.
- Bundey S (1996) Julia Bell MRCS LRCP FRCP (1879–1979). Steamboat Lady, Statistician and Geneticist. *Journal of Medical Biography*, **4**, 8–13.
- Cole TJ (2017) The remarkable life of Frances Wood. *Significance* `https://www.significancemagazine.com/science/563-the-remarkable-life-of-frances-wood`
- Cox DR (2001) *Biometrika*: the first 100 years. *Biometrika*, **88**, 3–11.

- Davis AEL (2017) Mathematical women – creating historical resources, *BSHM Bulletin: Journal of the British Society for the History of Mathematics*, **32**, 254–256. https://doi.org/10.1080/17498430.2017.1318248; see also https://mathshistory.st-andrews.ac.uk/Davis/

- Farewell V; Johnson T. (2010) Woods and Russell, Hill, and the emergence of medical statistics. *Statistics in Medicine*, **29**, 1459–1476.
- Farewell V; Johnson T; Gear R (2012) Hilda MaryWoods MBE,DSc,LRAM,FSS (1892–1971): reflections on a Fellow of the Royal Statistical Society. *JRSS–A*, **175**, 799–811.
- Farewell V; Johnson A (2012) The first British textbook of medical statistics. *Journal of the Royal Society of Medicine*, **105**, 446–448.
- Farewell V; Johnson T (2016) Major Greenwood (1880–1949): the biography. *Statistics in Medicine*, **35**, 5533–5535.

- Greenwood M; Yule GU (1920) An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple atacks of disease or of repeated accidents. *JRSS*, **83**, 255–279.
- Greenwood M; Thompson CM; Woods HM. (1925) Heights and weights of patients in mental hospitals. *Biometrika*, **17**, 142–158.
- Greenwood, M; Woods HM. (1919) A report on the incidence of industrial accidents upon individual with special reference to multiple accidents. *Reports of the Industrial Fatigue Research Board*, **4**. London.

- Hill ID (1984) Statistical Society of London – Royal Statistical Society: The first 100 years, 1834–1934. *JRSS–A* **147** 130–139.
- Pattison J et al (2013) Life expectancy: Past and future variations by gender in England & Wales. Institute of Actuaries, London.