CAMBRIDGE
UNIVERSITY PRESS

Article

Hemophilia B — Mutation Rates and Incidence — A Simulation Study

Alan E. Stark

Honorary Associate, School of Mathematics and Statistics FO7, University of Sydney, New South Wales, Australia

Abstract

The dynamics of rare X-linked recessive traits is explored by simulation. The model follows the prevalence of affected males and carrier females as separate but correlated variables. Different mutation rates and selection coefficients are introduced for males and females. A virtual population based on a published study of hemophilia B in the west of Scotland is followed at weekly intervals over many years. Speculative values of critical parameters to mimic the real population are proposed.

Keywords: Prevalence; X-linked; recessive; rare deleterious trait; hemophilia B; Scotland

(Received 29 January 2020; accepted 28 April 2020)

This article is concerned with the prevalence of rare X-linked recessive disorders that are maintained in the population by the opposing forces of mutation and selection. Hemophilia B (OMIM #306900 HEMB) is used as the model trait.

Curtis et al. (2015) convey an impression of the burden associated with hemophilia. They compared a cohort of young adults (age range 18–34 years) with comparable data from the general US population. They found that their cohort (103 with HEMA, 38 with HEMB) ranked worse on levels of educational achievement, employment, liver disease, joint damage, joint pain and physical health, compared to the general US population.

Kerr (1963, p. 359) divides the history of hemophilia into three parts: the first is the era of clinical observation and ineffectual treatment; the second is an age of increasingly effective medical management and better understanding of the underlying mechanisms; the third is 'yet to come'.

Ingram (1976) was the opening lecture to the Third European Regional Congress of the World Federation of Haemophilia, London, April 27, 1976, marking the Silver Jubilee of the British Haemophilia Society. Ingram summarizes the growth of knowledge of the disorder and the steps taken to treat those who suffer from it. He commends especially the contribution of Bulloch and Fildes (1911) that appeared in the series entitled *Treasury of Human Inheritance,* which was edited and supported by Karl Pearson. Ingram (p. 472) writes 'Bulloch and Fildes' great achievement was to define and stabilize the concept of haemophilia on the triad of symptoms, sex incidence and inheritance, and to rescue it from the confusion to which it had been subject'.

Mannucci and Tuddenham (2001, p. 1773) state that the modern management of hemophilia began in the 1970s with the use of plasma concentrates of coagulation factors. There were risks associated with these. They conclude that 'hemophilia is likely to be the first common severe genetic condition to be cured by gene therapy' (p. 1778).

Author for correspondence: Alan E. Stark, Email: alans@exemail.com.au

In an editorial, Ponder (2011) comments on the costs of protein therapy for hemophilia B and the welcome success of gene therapy reported in the same issue of the journal.

Verstraete and Vandenbroucke (1955, p. 1533) state 'In the last year haemophilia has been separated into different types – namely, the well-defined classical haemophilia A (antihaemophilic globulin deficiency) and Christmas disease or haemophilia B (Christmas factor deficiency, factor IX deficiency, plasma thromboplastin component deficiency)'.

Koller (1954) cites studies, including his own, contributing to this differentiation and gives a rudimentary diagram locating types A and B separately on the X chromosome. He states (p. 287) 'The Swiss bleeder family with the most extensively studied pedigree, the bleeders of Tenna, are affected with haemophilia B'.

Ledinot and Frézal (1961) acknowledge that hemophilia is due to two different factors, giving as evidence the fact that a mixture of sera from two hemophiliacs (A and B) leads to normal coagulation duration. These authors give a tentative map of the X chromosome (Figure 6, p. 104) showing Xg and deuteranopia on the short arm and G6PD, protanopia, HEMA and HEMB, well separated, on the long arm. They cite a number of other attempts to map the X chromosome.

Kalmus (1965, Chapter X, pp. 68–76) reviews early attempts to map the X chromosome using defective color vision. He reproduces a pedigree of hemophilia B and protanopia from Whittaker et al. (1962). He gives a 'hypothetical' map of the 'short arm' of the human X chromosome based on the studies which he reviews (p. 75). Subsequent corrections to this map show the difficulties encountered by early mapping attempts.

There are many examples of family trees, such as those of the descendants of Queen Victoria of England and those of Tenna, which show that a gene can be transmitted by carriers over several generations, so that what we study now may be the result of a mutation several generations ago.

For the intended study, information about mutation rates and selection coefficients are required. Ingram (1976, p. 469) gives the mutation rate as $1\text{–}4 \times 10^{-5}$, attributed to WHO (1972).

Cavalli-Sforza and Bodmer ([1971], Table 3.9, p. 96) give the mutation rate of HEMB as 5–10×10^{-6} and the birth frequency as 20–30×10^{-6}.

Ljung et al. ([2001], p. 81) state 'The ratio of male to female mutation rates was 5.3 and the overall mutation was 5.4×10^{-6} per gamete per generation'. In the simulation reported here, we take the male rate to be five times the female rate.

Darby et al. ([2007]) found that during the period 1977–1999, mortality in hemophiliacs exceeded mortality in the general population by a factor of 2.69.

Crew ([1947], p. 48) states 'On the average, haemophiliacs produce only about a quarter as many children as do normals. Affected males are usually removed by death before the age of twenty-five. They beget but few offspring therefore. For this reason alone female haemophiliacs must be rare since to be a haemophiliac a female must have had a haemophiliac for a father. In pooled pedigrees of haemophilia there is a significant deficiency of females and it would seem that the homozygous female commonly dies *in utero*'. No doubt the picture has changed greatly since this statement was printed.

Crew ([1947], p. 49) adds 'Queen Victoria was heterozygous for the haemophilic gene'. Subsequently, it was established that Victoria was a carrier of hemophilia B (Rogaev et al., [2009]).

The main concern of Connor et al. ([1985]) was to test a technique for identifying carriers of hemophilia B using an intragenic genomic probe. They give data that we use to construct a virtual population: 52 males with hemophilia B were ascertained with stated prevalence 1/26870; this implies a male population of approximately 1.4 million and a combined population of 2.8 million. This is the number used to simulate the actual population. The details of the model are followed by a realization and discussion.

The Model

The equations expressing the balance between mutation and selection are

$$Xrb_X/4 + Arb_A/2 + Srm = Xrd_X \tag{1}$$

and

$$Xrb_X/4 + Srm/2 = Ard_A. \tag{2}$$

S is the total size of the population, r is the reproduction rate, X is the number of carrier females, A is the number of hemophiliacs, b_X is the rate of reproduction in carrier females relative to r, b_A is the corresponding parameter in hemophiliacs, d_X is the relative death rate in carriers, d_A is the corresponding rate in hemophiliacs and m is the mutation rate.

Matching these gives the following values in equilibrium:

$$\tilde{X} = \frac{2Sm(4d_A + b_A)}{2d_A(4d_X - b_X) - b_X b_A} \tag{3}$$

and

$$\tilde{A} = \frac{Sm(4d_X + b_X)}{2d_A(4d_X - b_X) - b_X b_A}. \tag{4}$$

In equilibrium, the ratio of $\tilde{X}$ to $\tilde{A}$ is

$$\frac{2(4d_A + b_A)}{4d_X + b_X}. \tag{5}$$

If different rates of mutation are specified for eggs and sperm, specifically m and s, equation (1) is replaced by

$$Xrb_X/4 + Arb_A/2 + Srm/2 + Srs/2 = Xrd_X \tag{6}$$

while no change is required for (2). The equilibrium solutions are

$$\tilde{X}_s = \frac{2S(2d_A(m+s) + b_A m)}{2d_A(4d_X - b_X) - b_X b_A} \tag{7}$$

and

$$\tilde{A}_s = \frac{S(4d_X m + b_X s)}{2d_A(4d_X - b_X) - b_X b_A}. \tag{8}$$

The ratio of $\tilde{X}_s$ to $\tilde{A}_s$ is

$$\frac{2(2d_A(m+s) + b_A m)}{4d_X m + b_X s}. \tag{9}$$

The proportion of cases from new mutation in equilibrium is

$$p_m = \frac{Srm/2}{\tilde{A}_s r d_A} = \frac{Sm}{2\tilde{A}_s d_A}. \tag{10}$$

This equation suggests how the mutation rate in eggs might be estimated directly as:

$$\hat{m} = \frac{2\hat{p}_m A d_A}{S}. \tag{11}$$

Exploiting (2) and (6) suggests a way of estimating as:

$$\hat{s} = \frac{X(4d_X - b_X) - 2(Ab_A + S\hat{m})}{2S}. \tag{12}$$

In practice, the value of X will not be known accurately and estimates of d_X and so on are required.

Under the model, in equilibrium, the incidence (i.e., the proportion of male live births) of HEMB in males is

$$\tilde{I}_A = \frac{\tilde{A}_s r d_A}{Sr/2} = \frac{2\tilde{A}_s d_A}{S} = \frac{2d_A(4d_X m + b_X s)}{2d_A(4d_X - b_X) - b_X b_A}. \tag{13}$$

The frequency of the trait gene in equilibrium $2(\tilde{A}_s + \tilde{X}_s)/(3S)$ is

$$\tilde{q} = \frac{2(2m(b_A + 2(d_X + d_A)) + s(b_X + 4d_A))}{3(2d_A(4d_X - b_X) - b_X b_A)}. \tag{14}$$

The simulation begins by nominating an arbitrary number of cases A_0 and arbitrary number of carriers X_0, with reason similar to equilibrium values, but not necessarily so. The process proceeds week by week jointly for cases and carriers. The reproduction rate r

is given the value 15/1000 per year. A different arbitrary value produces the same predictions.

A random number of mutants from eggs is drawn from the Poisson distribution with mean $Srwm/2$, where w is the fraction of a year taken up by a week. The number of cases inheriting the trait from the current number of carriers X is drawn from the binomial distribution with number of 'trials' X and probability of 'success' $rwb_X/4$. From the current number of cases, a random number of deaths is drawn from the binomial distribution with number of trials A and probability of success rwd_A. The number of cases is updated by adding the first two and deducting the third from the existing number of cases.

A random number of mutants from eggs is drawn from the Poisson distribution with mean $Srwm/2$ and a random number with mean $Srws/2$ from sperm. A random number is drawn from the binomial distribution with number of trials A and probability of success $rwb_A/2$. A random number is drawn from the binomial distribution with number of trials X and probability of success $rwb_X/4$. From the current number of carriers X, a random number of deaths is drawn from the binomial distribution with number of trials X and probability of success rwd_X. The number of carriers is updated by adding the first four and deducting the last random number. The process is repeated for the desired number of weeks. In the process just described, any possible coincidences are ignored on the grounds that the amount of change in a week is small if it is not zero.

Realization of the Model

Connor et al. (1985) touch on the difficulties facing genetic counselors, which is their main preoccupation and so do not go into the question of rates that are represented by the parameters in equations (1), (2) and (6). Here we simply attempt to match their 52 prevalent cases and seek parameters of mutation and selection elsewhere. Using data from papers cited earlier, we take $S = 2,800,000$, $m = 12 \times 10^{-6}$, $s = 60 \times 10^{-6}$, $b_X = 7/4$, $b_A = \frac{1}{2}$, $d_X = 1$ and $d_A = 2$. The rounded predicted number with the trait at equilibrium is 52 and the number of carriers is 203.

A single realization of the model was effected starting with 200 carriers and 50 with the trait. The population was followed for 1000 years, a year being 52 weeks of 7 days. Figure 1 depicts the numbers with the trait and Figure 2 the numbers of carriers over the whole period. While there are only small changes over say 3 years, there can be quite large changes over long intervals. There is low correlation between the number with the trait and the number of carriers in the short run, but moderate cross-correlation in the long run.

The median number of individuals with the trait over the whole interval was 51, of carriers 194, the minima 35 and 161, and the maxima 78 and 248, respectively. The changes from week to week are summarized in Table 1, showing that changes of more than one in either the number with the trait or carriers, or both, are infrequent. While it may seem absurd to consider a phenomenon over 1000 years in a rapidly changing world, we may not know accurately the current number of carriers.
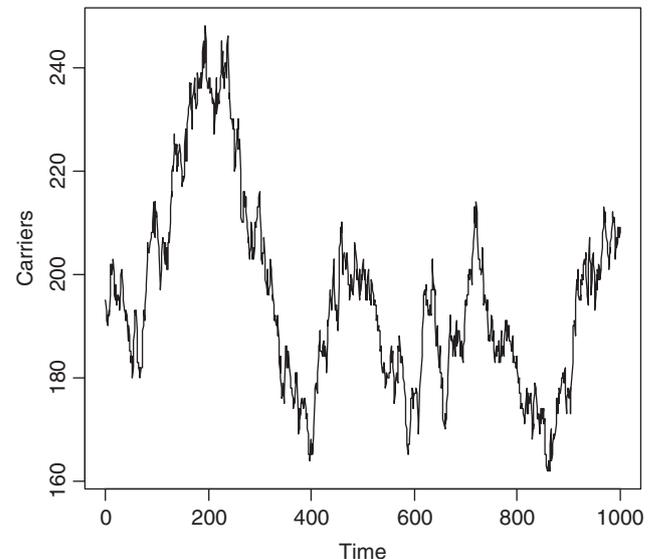
The expected number of hemophiliacs dying each year is Ard_A, so the number of years required to replace the existing ones is $1/(rd_A)$. This is approximately 33 and serves as an estimate of average lifespan. A similar argument applying to carriers yields the estimate 67, and the ratio of lifespans $d_X/d_A = 1/2$, in this case. Obviously, the reliability of this argument depends on the factual basis of the parameters.

Table 1. Simulated frequencies of changes in numbers of carriers and patients from week to week over 1000 years

Carriers	Patients				
	−2	−1	0	1	2
−3	0	0	2	0	0
−2	0	2	84	1	0
−1	0	76	2454	69	1
0	14	1329	43,922	1320	20
1	3	71	2478	71	1
2	0	1	75	2	0
3	0	0	3	0	0
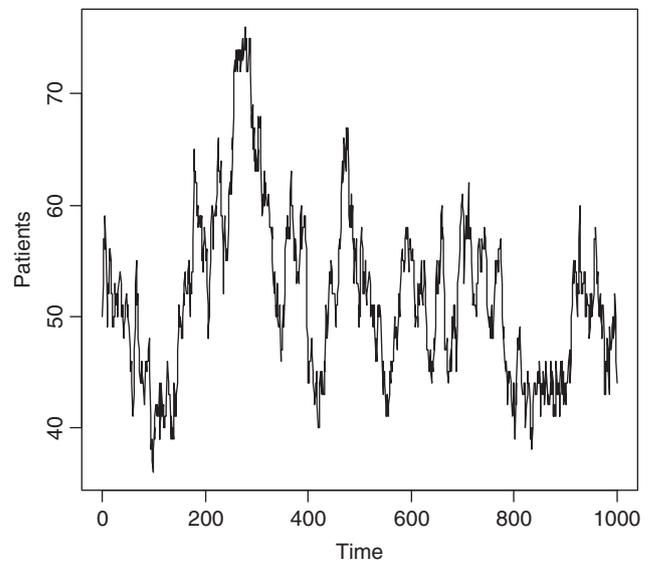
Fig. 1. Simulated numbers of patients at yearly intervals ($S = 2,800,000$, $m = 12 \times 10^{-6}$, $s = 60 \times 10^{-6}$, $b_X = 7/4$, $b_A = \frac{1}{2}$, $d_X = 1$ and $d_A = 2$).

Fig. 2. Simulated numbers of carrier females at yearly intervals ($S = 2,800,000$, $m = 12 \times 10^{-6}$, $s = 60 \times 10^{-6}$, $b_X = 7/4$, $b_A = \frac{1}{2}$, $d_X = 1$ and $d_A = 2$).

The expected incidence of hemophilia B in males, from equation (13), using the speculative parameters, is 1/13,276.

Discussion

Mannucci and Tuddenham (2001) give summary statistics for both A and B types of hemophilia. They state that about 30% of cases arise with no family history. The equilibrium prediction from equation (10) for type B is about 16%. These authors state that the incidence of A is 1 in 5000 and of B is 1 in 30,000.

Kling et al. (1992) state that of 45 hemophilia B patients registered at the hemophilia center in Malmö, Sweden, 24 were the sole members of their families to be affected. In 13 of the 24, ascendant relatives were available for study. Ten of these inherited the defect from mothers and three, that is roughly one-quarter, had de novo mutations. All the 10 carrier mothers had *de novo* mutations. Six of the 10 carrier mothers were informative as to the origin of the mutation, which were all paternal. These authors report 'the relatively high turnover of the haemophilia-B gene pool' (p. 144).

Ljung et al. (2001) give a succinct summary of the characteristics of hemophilia B based on Swedish data. They divide cases into 'severe', 'moderate' and 'mild'. The range and complexity of their findings show that a model such as the one presented here presents a very simplified picture.

Mannucci and Tuddenham (2001) discuss the use of plasma concentrates of coagulation factors and transmission of hepatitis virus and HIV. Medical intervention, while desirable for families, creates problems for epidemiologists.

Crew (1947, p. 49) contains some population genetics theory: 'If the frequency of haemophilic males among the male population is p and the frequency of normal males is q, where $(p + q) = 1$, then with random mating, the frequency of haemophiliacs, carriers and normals among the female population is p^2: $2pq$: q^2'. The derivation is based on the assumption that the frequency of the hemophilic gene is the same in males as females. In the notation used here this is equivalent to asserting $2A = X$, which is not generally the case in the model population.

Applying equation (14) predicts HEMB gene frequency to be 6.08×10^{-5} and the heterozygote, that is carrier number, from the Hardy–Weinberg distribution evoked by Crew, in the virtual population to be 170. Figure 2 suggests that this is too low.

References

Bulloch, W., & Fildes, P. (1911). *Treasury of human inheritance*, Parts V and VI, Section XIVa, Haemophilia. Francis Galton Laboratory for National Eugenics, University of London. London: Dulau and Co.

Cavalli-Sforza, L. L., & Bodmer, W. F. (1971). *The genetics of human populations.* San Francisco, CA: W. H. Freeman and Company.

Connor, J. M., Pettigrew, A. F., Hann, I. M., Forbes, C. D., Lowe, G. D. O., & Affara, N. A. (1985). Application of an intragenic genomic probe to genetic counselling for haemophilia B in the west of Scotland. *Journal of Medical Genetics*, 22, 441–446.

Crew, F. A. E. (1947). *Genetics in relation to clinical medicine.* Edinburgh: Oliver and Boyd.

Curtis, R., Baker, J., Riske, B., Ullman, M., Niu, X., Norton, K., & Nichol, M. B. (2015). Young adults with hemophilia in the US: Demographics, comorbidities, and health status. *American Journal of Hematology*, 90, S11–S16.

Darby, S. C., Kan, S. W., Spooner, R. J., Giangrande, P. L. F., Hill, F. G. H., Hay, C. R. M., & Williams, M. (2007). Mortality rates, life expectancy, and causes of death in people with haemophilia A or B in the United Kingdom who were not infected with HIV. *Blood*, 110, 815–825.

Ingram, G. I. C. (1976). The history of haemophilia. *Journal of Clinical Pathology*, 29, 469–479.

Kalmus, H. (1965). *Diagnosis and genetics of defective colour vision.* Oxford: Pergamon Press.

Kerr, C. B. (1963). The fortunes of haemophiliacs in the nineteenth century. *Medical History*, 7, 359–370.

Kling, S., Ljung, R., Sjörin, E., Montandon, J., Green, P., Giannelli, F., & Nilsson, I. M. (1992). Origin of mutation in sporadic cases of haemophilia-B. *European Journal of Haematology*, 48, 142–145.

Koller, F. (1954). Is hemophilia a nosologic entity? *Blood*, 9, 286–290.

Ledinot, E., & Frézal, J. (1961). Ébauche de carte factorielle du chromosome X. *Annales de Génétique*, 7, 101–105.

Ljung, R., Petrini, P., Tengborn, L., & Sjörin, E. (2001). Haemophilia B mutations in Sweden: A population-based study of mutational heterogeneity. *British Journal of Haematology*, 113, 81–86.

Mannucci, P. M., & Tuddenham, E. G. D. (2001). The hemophilias from royal genes to gene therapy. *New England Journal of Medicine*, 344, 1773–1779.

Ponder, K. P. (2011). Merry Christmas for patients with Hemophilia B. *New England Journal of Medicine*, 365, 2424–2425.

Rogaev, E. I., Grigorenko, A. P., Faskhutdinova, G., Kittler, E. L. W., & Moliaka, Y. K. (2009). Genotype analysis identifies the cause of the 'Royal Disease'. *Science*, 326, 817.

Verstraete, M., & Vandenbroucke, J. (1955). Combined antihaemophilic globulin and Christmas factor deficiency in haemophilia. *British Medical Journal*, 2, 1533–1535.

Whittaker, D. L., Copland, D. L., & Graham, J. B. (1962). Linkage of color blindness to hemophilias A and B. *American Journal of Human Genetics*, 14, 149–158.