TRANSACTIONS OF SOCIETY OF ACTUARIES
1992 VOL. 44

THE BIRTHDAY RULE AND THE DIFFERENCE IN SPOUSES' AGES

BERTRAM M. KESTENBAUM

ABSTRACT

The "birthday rule" is often used to determine whether the father's insurer or the mother's insurer is liable for a child's medical bills. Implicit in the acceptance of this criterion is the assumption that the likelihood that a husband's birthday falls earlier in the year and the likelihood that a wife's birthday falls earlier are about 50–50.

The validity of this assumption is an empirical question. This paper demonstrates how the key determinant is the distribution of the difference in spouses' ages. Data are presented showing that the assumption *is* valid for the U.S.

INTRODUCTION

The "birthday rule" often is the criterion for determining who is liable for a child's medical bills, the father's insurer or the mother's insurer. According to this rule, liability rests with the insurer of the parent whose birthday falls earlier in the year. This paper considers whether this criterion is equitable for insurers operating in predominantly male or predominantly female industries.

An equivalent question is: Is the birthday of a randomly selected woman equally likely to precede or to follow her husband's? The answer is not obvious, because the husband is not an independent random selection from the male population. And although this question at first appears devoid of sociological substance, we soon show this is not the case.

Then we demonstrate how this question can be viewed in a more general framework, and we recognize its similarity to a familiar problem in the evaluation of census data. Afterwards we present the evidence that for the U.S. the question can be answered in the affirmative.

THE KEY VARIABLE

Upon reflection, the difference in spouses' ages emerges as the variable upon which this question turns. The role of this sociological parameter can be easily appreciated from a simple example. Consider a hypothetical country in which one-half of the women marry men one month older, another one-third marry men two months older, and the remaining one-sixth marry men three months older. Also, men's birthdays in this country are uniformly distributed throughout the year. Then the probability that the wife's birthday will fall later in the year than her husband's in this imaginary country is 0.86, the result of the calculation.

$$(1/2) \times (11/12) + (1/3) \times (10/12) + (1/6) \times (9/12).$$

In general, the probability that the wife's birthday occurs later in the year is determined by the distribution of the difference in spouses' ages in the following way, so long as the distribution of birthdays throughout the year among the male population is near uniform. If age is measured in months, consider the husband's age less the wife's age, *modulo 12*. Then the probability that the wife's birthday is later can be approximated by the sum of 12 products, where the i-th ($i = 1, 2, \ldots, 11$) is the product of the probability of i in the modulo-12 distribution by the factor $(12-i)/12$, and the 12th is the product of the probability of 0 by the factor 1/2. Values for i of (1, 2, 3, 4, and 5) are associated with a greater likelihood that the husband's birthday comes earlier, while values for i of (7, 8, 9, 10, and 11) are associated with a greater likelihood that the wife's birthday comes earlier. The probability in question will equal 0.5 if the modulo-12 distribution is uniform; otherwise, it may or may not.

THE MODULO-K FRAMEWORK

This approach to focusing on the impact of a nonuniform modulo-k distribution was inspired by Myers' [2] discussion of the measurement of the phenomenon of "digital preference" in census and other data (that is, the tendency of persons to misreport their ages as ages ending in certain digits, such as 0 or 5). A simple measure of the extent of this phenomenon in some data set is based on a comparison of the relative frequency distribution for the units digit of reported age with a uniform distribution having 10 percent in each class. Myers recognized that this measure is flawed: in most real populations, as well as in all life table populations, the actual frequency distribution decreases with age. Even in the absence of any digital preference, more persons will be reported as age 0 than as age 9, more as age 10 than as age 19, more as age 20 than as age 29, and so on.

In our terminology, Myers' insight was that the modulo-10 distribution of true ages in the population is not uniform. Rather, it is weighted towards the lower end. Therefore, the use of a uniform distribution as the standard for evaluating the accuracy of age reporting in a census or other data collection is not technically correct.

Restating Myers' observation in a more abstract form suggests how it might be generalized. Myers observed that if $F(X,0)$, the frequency distribution of X beginning with $x=0$ and proceeding through the positive integers, is a decreasing function of X, then $G(X,0,10)$, the modulo-10 frequency distribution of X, decreases throughout its domain of the 10 integers (0, 1, ..., 9). Now, neither the initial value of 0 nor the modulus of 10 is irreplaceable, hence the following more general result. If $F(X,a)$, the frequency distribution of X beginning with $x=a$ and proceeding through the higher integers, is a decreasing function of X, then $G(X,a,k)$, the modulo-k frequency distribution of X, decreases throughout its domain of the k integers taken in order $\{a$ modulo-k, $(a+1)$ modulo-k, ..., $(a+k-1)$ modulo-$k\}$. For convenience we denote "n modulo-k" as $n|k$.

This formulation is appropriate even if the initial value is a negative integer. Recall that for negative n, the modulo-k value is defined to be $k-[(-n)|k]$. For example, -15 modulo-12 is equal to 9.

The following more general statement would cover the case of an increasing distribution as well. If $F(X,a)$, the frequency distribution of X beginning/ending with $x=a$ and proceeding through the higher/lower integers, is a decreasing/increasing function of X, then $G(X,a,k)$, the modulo-k frequency distribution of X, achieves its maximum at $a|k$ and otherwise decreases/increases through the $(k-1)$ integers taken in order $\{(a+1)|k, (a+2)|k, ..., (a+k-1)|k\}$.

Now, how can we characterize the modulo-k distribution for a frequency distribution that is first increasing up to its modal value at a and then decreasing afterwards? Over the subdomain where X is greater than or equal to a, the modulo-k distribution decreases through the ordered set $\{a|k, (a+1)|k, ..., (a+k-1)|k\}$, while over the subdomain where X is less than a, it increases over the same set. Thus the tendency is towards balancing out; but, intuitively, the actual distribution will vary from the uniform distribution in magnitude and direction as determined by the comparative sizes of the total frequencies in the two subdomains and the rates of decrease/increase within each.

In terms of the frequency distribution of the (signed) difference in age between husband and wife, measured in months, if the frequencies increase up to a modal difference of a and then decrease beyond that point, the shape of the net modulo-12 distribution depends on the comparative numbers of couples where the husband is at least a months older than his wife and where the husband is less than a months older and on the slopes of the frequency functions within each of these two groups of couples.

U.S. DATA

Each country's age-at-marriage pattern determines whether the probability in that country is approximately 50 percent that the birthday of a randomly selected woman follows her husband's in the calendar year. In the U.S. this *is*, in fact, the case. The National Center for Health Statistics prepared for us a cross-tabulation of month of birth of groom with month of birth of bride in a sample of over 100,000 marriages taking place in 1987 in the 38 states that report this information to the National Center. According to this tabulation, the groom had the earlier month of birth 42.5 percent of the time, and the bride, 42.6 percent of the time. In 7.9 percent of these marriages bride and groom were born in the same month, and the data are incomplete for the remaining 7.0 percent of marriages.

Because these data pertain only to current marriage patterns and because the National Center could not as easily provide information on the more substantive issue of the frequency distribution of the difference in spouses' ages, we sought other data sources. One source is a 1-in-100 sample of administrative records for couples receiving Social Security benefits in December 1990 based on the same work record, one member of the couple as the primary beneficiary and the other as an auxiliary beneficiary. A large majority of these couples are elderly, so the emerging picture reflects for the most part the marriage patterns of several decades ago.

In a comparison of the *exact* birthdays of husband and wife for the 43,511 couples in the sample, the husband's is earlier 49.7 percent of the time. The wife's is earlier 49.9 percent of the time, and husband and wife have the same birthday 0.3 percent of the time.

The modulo-12 relative frequency distribution of husband's age minus wife's age, in months, for the Social Security beneficiary couples is shown in the left-hand panel of Table 1. Consistent with the results given in the prior paragraph, the deviation from a pattern of uniformity is slight.

TABLE 1

RELATIVE FREQUENCY DISTRIBUTIONS OF DIFFERENCE IN SPOUSES' AGES, IN MONTHS, MODULO-12

Husband's Age Minus Wife's Age, Modulo-12	Beneficiary Couples, Dec. 1990 (1% Sample of Administrative Records)	Population, March 1973 (Current Population Survey)
Total	100.0%	100.0%
0	8.7%	9.4%
1	8.2	8.2
2	8.3	8.2
3	8.5	8.4
4	8.3	8.5
5	8.5	8.4
6	8.3	8.3
7	8.5	8.1
8	8.4	8.2
9	8.1	8.3
10	7.9	8.1
11	8.4	8.1

If month-of-birth information were present in the records of decennial census public-use files or Current Population Survey public-use files, a cross-sectional perspective on age-at-marriage patterns could be obtained. In general, it is not, but an exception is the Exact Match file (Kilss and Scheuren [1]), constructed by supplementing the March 1973 Current Population Survey with administrative data.

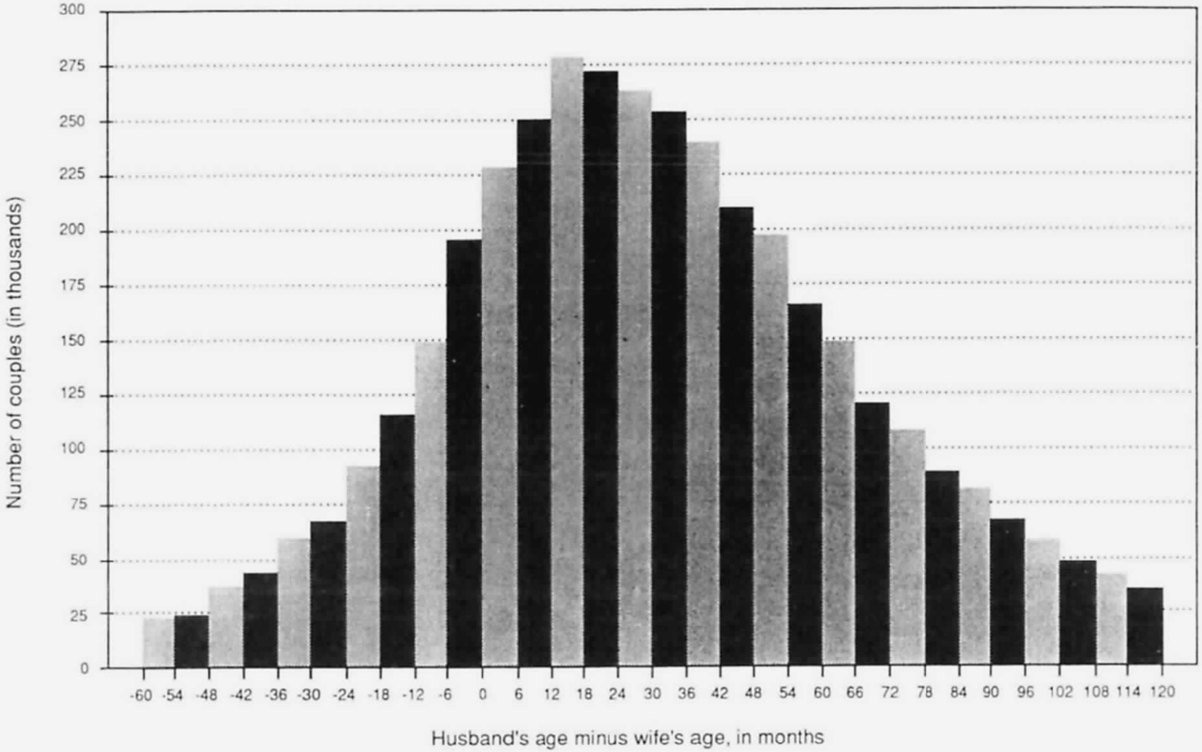
In a sample of 29,058 couples in this file, the husband's month of birth was earlier 45.5 percent of the time, the wife's 45.1 percent of the time, and there was no difference 9.4 percent of the time. The modulo-12 distribution in this sample is given in the right-hand panel of Table 1.

The modulo-k distribution of differences in spouses' ages reflects an underlying frequency distribution that first increases up to a modal value and then decreases. This latter distribution for Social Security beneficiary couples is depicted in Figure 1. To make the information more manageable, the interval width is 6 months and the extremes of the distribution are not shown. The presentation in terms of 6-month intervals is meaningful because the husband's birthday is much more likely to be earlier if the difference in ages in months is 0–5, 12–17, 24–29, and so on, while the wife's is much more likely to be earlier if the difference in months is 6–11, 18–23, 30–35, and so on.

The mode for Social Security beneficiary couples is the 12 month-17 month interval. In the subdomain beginning with this interval, the husband's

FIGURE 1

DISTRIBUTION OF SOCIAL SECURITY BENEFICIARY COUPLES BY DIFFERENCE IN SPOUSES' AGES, DECEMBER 1990 (BASED ON 1-IN-100 SAMPLE)



birthday is likely to be the earlier of the two; in the subdomain in which the husband is younger than his wife or is older by less than 12 months, the wife's birthday is likely to be the earlier. Although many more marriages are in the first subdomain than in the second, the pattern of decrease in the first is also much more gradual than the pattern of increase in the second. The net modulo-12 distribution is not markedly different from the uniform.

CONCLUSION

The distribution of the difference in age between husband and wife determines whether the two have an equal probability of being born later in the calendar year, and hence the equity of the "birthday rule." We see that the probabilities are equal in the U.S., and therefore the birthday rule puts at a disadvantage neither the insurer whose business is concentrated in predominantly male industries nor the insurer whose business is concentrated in predominantly female industries.

REFERENCES

1. KILSS, B., AND SCHEUREN, F. "The 1973 CPS-IRS-SSA Exact Match Study," *Social Security Bulletin* 41 (October 1978): 14–22.
2. MYERS, ROBERT J. "Errors and Bias in the Reporting of Ages in Census Data," *TASA* 41 (1940): 395–415.

DISCUSSION OF PRECEDING PAPER

CHARLES S. FUHRER:

Mr. Kestenbaum is to be commended for bringing to our attention a rating problem that results from the birthday rule. The birthday rule is part of a coordination of benefits (COB) provision that states whether the father's or the mother's insurer will be primary or secondary when the children's medical bills are covered by both insurers. Paying primary means that the insurer pays its standard benefits without regard to the other coverage. Paying secondary means reducing the benefits based on the other insurer's payments according to the formula in the COB provision. The rating problem is that for a given group each insurer may not actually be primary, on exactly half of the children for which both spouses are covered by different insurers. Note that the author has stated the problem as one of equitableness. He also has implied incorrectly that the insurer of the spouse with the later birthday is not liable for any of the children's medical bills if the other spouse has family coverage with a different insurer.

Unfortunately, Mr. Kestenbaum has addressed only one part of the problem. The problem that he has addressed is whether the birthday rule is equitable for insurers operating in a predominantly male or female market. A more significant problem would be how much error is introduced in the rating process by ignoring whether each of the children covered in a group is primary or secondary. Most group insurers completely ignore the actual birth dates of the individual insureds. Of course, the birth dates are often unavailable or expensive to tabulate. Insurers usually calculate a size of claim distribution for rating children that includes both primary and secondary claims. A better way would be to form a separate size of claim distribution for primary and secondary children's coverage and then rate each dependent coverage based on the actual birth date of the employees. Let the rate for primary be R_p, the rate for secondary R_s, and the probability of being primary p. Then the rate would be equal to $pR_p+(1-p)R_s$. That is, children of employees with birthdays early in the year would be rated with mostly the primary rate and those with birthdays late in the year mostly the secondary rate. It would have been useful if the author had developed some estimate of the size of the error (or the extra variance) by size of group that results from ignoring the actual birthdays.

I would like clarification of several points. What is the approximation that is being used in the last paragraph of the section entitled "The Key Variable"? What is the intent of the section on the modulo-k framework? What does the Myers discussion have to do with the birthday rule?

Here is my version of the second and third sections of the paper:

Let the age of one spouse (say, the wife's) be the random variable X and the age of the husband be the random variable Y, both measured on December 31. Let $f(z|x)$ be the conditional density of $Z=Y-X$ given $X=x$. Define $[x]$ as the greatest integer smaller or equal to x. Then we have the probability of male priority, given $X = x$, as:

$$\begin{aligned} Pr\{X - [X] < Y - [Y] | X = x\} &= \sum_{i=-\infty}^{\infty} Pr\{X - [X] < Y - [Y] \\ &\qquad \text{and } [Y] = [X] + i \mid X = x\} \\ &= \sum_{i=-\infty}^{\infty} Pr\{x - [x] < Y - [x] - i \\ &\qquad \text{and } [x] + i \le Y < [x] + i + 1\} \\ &= \sum_{i=-\infty}^{\infty} Pr\{x + i < Y < [x] + i + 1\} \\ &= \sum_{i=-\infty}^{\infty} \int_{i}^{[x]-x+i+1} f(z|x)dz. \end{aligned}$$

Now let $X=[X]+W$ and assume W is uniform on the interval (0,1) and $[X]$ and W are independent. Further, let $g(x)$ be the density function of X and $p_j=Pr\{[X]=j\}$. Then unconditionally, the probability of male priority is

$$\begin{aligned} &Pr\{X - [X] < Y - [Y]\} \\ &\qquad = \int_0^{\infty} Pr\{X - [X] < Y - [Y] | X = x\} g(x)dx \\ &\qquad = \int_0^{\infty} \sum_{i=-\infty}^{\infty} \int_{i}^{[x]-x+i+1} f(z|x)dz\,g(x)dx. \end{aligned}$$

Now split the outer integral into unit intervals. In the outer integral, make the substitution, $w=x-j$. Note that on the interval $[j, j+1)$, $x=j+w$. Further assume that Z is independent of X. Then reverse the order of integration to obtain:

$$Pr\{X - [X] < Y - [Y]\} = \sum_{j=0}^{\infty} \int_{j}^{j+1} \sum_{i=-\infty}^{\infty} \int_{i}^{[x]-x+i+1} f(z)p_j dzdx.$$

$$= \sum_{j=0}^{\infty} p_j \int_{j}^{j+1} \sum_{i=-\infty}^{\infty} \int_{i}^{[x]-x+i+1} f(z)dzdx.$$

$$= \left(\sum_{j=0}^{\infty} p_j\right)\left(\sum_{i=-\infty}^{\infty} \int_{0}^{1} \int_{i}^{-w+i+1} f(z)dzdw\right)$$

The first sum is 1, as it is just the sum of probabilities.

$$= \sum_{i=-\infty}^{\infty} \int_{i}^{i+1} \int_{0}^{1-z+i} f(z)dwdz$$

$$= \sum_{i=-\infty}^{\infty} \int_{i}^{i+1} (1 - z + i)f(z)dz$$

$$= \int_{0}^{\infty} (1 - z + [z])f(z)dz$$

$$= 1 - E(Z) + E([Z]).$$

Therefore:

$$p_f(Z) = Pr\,\{fem\text{-}pri\} = Pr\{Y - [Y] < X - [X]\} = E(Z) - E([Z]).$$

Thus the probability that a male (or female) has priority depends only on the distribution of the difference of the ages (given the assumptions above).

For certain classes of distributions of Z, the value of $p_f(Z)$ goes to 1/2 as the variance increases. The unimodal distributions are such a class. A distribution with density f is unimodal with mode m if $f(x) \leq f(m)$ and $x < y \leq m$ or $x > y \geq m$ implies $f(x) \leq f(y)$, for all x and y. For a scale parameter r, let Z_r have the density $f(m + (x-m)/r)/r$. Then Var $(Z_r) = r^2 Var(Z)$ and:

$$\lim_{r \to \infty} p_f(Z_r) = 1/2.$$

The normal distribution is unimodal with scale parameter σ.

Here are some values of p_f for Z having the normal distribution with parameters μ and σ:

μ	σ 0.05	0.10	0.25	0.50
0.00	0.500	0.500	0.500	0.500
0.05	0.050	0.359	0.471	0.499
0.10	0.100	0.259	0.444	0.499
0.15	0.150	0.217	0.424	0.498
0.20	0.200	0.223	0.411	0.498
0.25	0.250	0.256	0.407	0.498
0.30	0.300	0.301	0.413	0.498
0.35	0.350	0.350	0.426	0.498
0.40	0.400	0.400	0.447	0.499
0.45	0.450	0.450	0.472	0.499
0.50	0.500	0.500	0.500	0.500
0.55	0.550	0.550	0.528	0.501
0.60	0.600	0.600	0.553	0.501
0.65	0.650	0.650	0.574	0.502
0.70	0.700	0.699	0.587	0.502
0.75	0.750	0.744	0.593	0.502
0.80	0.800	0.777	0.589	0.502
0.85	0.849	0.783	0.576	0.502
0.90	0.877	0.741	0.556	0.501
0.95	0.791	0.641	0.529	0.501
1.00	0.500	0.500	0.500	0.500

In Figure 1, the author presents the distribution of the difference in spouses ages. I wish he had stated the sample mean and variance for these data. I was able to estimate from the figure that the mean was 2.45 and the standard deviation over 3. The shape of the distribution appears close to normal. It is bell-shaped with only a little skewness. The p_f for a Z that is normally distributed with $\sigma \geq 1.0$ would be equal to 0.5 ± 0.00000001. Thus, it is not surprising that the comparison of the exact birthdays in the sample gave values close to 0.5.

(AUTHOR'S REVIEW OF DISCUSSION)

BERTRAM M. KESTENBAUM:

Because my paper was, I thought, quite straightforward, I was surprised to receive a discussion that reflects difficulties in comprehension of the material. In particular, the thesis of the paper is that the distribution of the difference in spouses' ages *might* be characterized by an irregular pattern that could have a negative impact on the equity of the "birthday rule"; the empirical finding is that this is not the case. Mr. Fuhrer's calculations of the smallness of the departure from equity for many *normal* distributions is therefore irrelevant.

Because of the questions raised, I feel compelled to sketch here the paper's line of argument. I showed, first by an example and then more generally, that the distribution of the difference in spouses' ages is the key factor in determining the overall likelihood that a husband's birthday falls earlier in the calendar year than his wife's. I then pointed out that if the difference in ages is y years and m months, only the value of m is relevant, not the value of y. This led to a useful restatement of the problem in terms of the difference in ages *modulo-12.* (The equivalence to Mr. Fuhrer's $Z-[Z]$ function is obvious.) I noted a precedent for this framework in the actuarial literature, citing Myers' classic paper on digit preference in age reporting in censuses and surveys, and showed how our work is a generalization of his.

Mr. Fuhrer correctly points out that under the coordination-of-benefits provisions, the secondary insurer often has reduced liability, rather than none at all, but this is of course a technical, rather than a substantive, correction. Mr. Fuhrer states that he would have preferred an analysis of how a company's rating of secondary claims would be affected were the distribution of its clients' birthdays specifically included in the rating exercise; the "bottom line" is that there would be little effect, even if the clientele were predominantly of one sex.

I am uncertain what value to place on Mr. Fuhrer's complex exercise to confirm what can be deduced promptly from probability fundamentals.

Finally, I wonder whether the *Transactions* readership shares my view of the role of discussions in a professional journal or at a professional meeting, that is, to provide constructive criticism and direction for further work in a spirit of respect and camaraderie.