Research

On the number of founding germ cells in humans
Chang-Jiang Zheng*1, E Georg Luebeck2, Breck Byers3 and Suresh H Moolgavkar2

Address: 1Department of Occupational and Environmental Medicine, Regions Hospital, University of Minnesota, 640 Jackson Street, Saint Paul, MN 55101, USA, 2Division of Public Health Sciences Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA and 3Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

Email: Chang-Jiang Zheng* - zhen0075@umn.edu; E Georg Luebeck - gluebeck@fhcrc.org; Breck Byers - byers@gs.washington.edu; Suresh H Moolgavkar - smoolgav@fhcrc.org
* Corresponding author

Abstract

Background: The number of founding germ cells (FGCs) in mammals is of fundamental significance to the fidelity of gene transmission between generations, but estimates from various methods vary widely. In this paper we obtain a new estimate for the value in humans by using a mathematical model of germ cell development that depends on available oocyte counts for adult women.

Results: The germline-development model derives from the assumption that oogonial proliferation in the embryonic stage starts with a founding cells at \( t = 0 \) and that the subsequent proliferation can be defined as a simple stochastic birth process. It follows that the population size \( X(t) \) at the end of germline expansion (around the 5th month of pregnancy in humans; \( t = 0.42 \) years) is a random variable with a negative binomial distribution. A formula based on the expectation and variance of this random variable yields a moment-based estimate of \( a \) that is insensitive to the progressive reduction in oocyte numbers due to their utilization and apoptosis at later stages of life. In addition, we describe an algorithm for computing the maximum likelihood estimation of the FGC population size (\( a \)), as well as the rates of oogonial division and loss to apoptosis. Utilizing both of these approaches to evaluate available oocyte-counting data, we have obtained an estimate of \( a \approx 2 – 3 \) for Homo sapiens.

Conclusion: The estimated number of founding germ cells in humans corresponds well with values previously derived from chimerical or mosaic mouse data. These findings suggest that the large variation in oocyte numbers between individual women is consistent with a smaller founding germ cell population size than has been estimated by cytological analyses.

1. Introduction

Despite great strides in our understanding of the genetic regulation of germ cell determination in recent years [1], the size of the founding germ cell population in humans remains obscure. Due to this uncertainty, it is difficult in a clinical environment to estimate the probability that a mutant allele known to be mosaic in the somatic tissues of a parent will be transmitted to offspring. Even in the mouse, where experimental approaches are feasible, the number of founding germ cells (FGCs) has proven
difficult to establish. Cytochemical methods have suggested FGC numbers varying from 45 cells [2] to 193 cells [3]. On the other hand, genetic analysis of artificially generated chimerical cellular populations in the mouse indicate that there are only 2 to 9 cells that actually contribute to the germ cell population [4-10]. In this communication, we derive a new method that is applicable to both humans and laboratory animals. This approach exploits a “founder effect” phenomenon that has previously been shown to be amenable to mathematical analysis [11,12]. Specifically, such analysis has shown that, if a population is descended from a small set of ancestral founders, the population size should exhibit substantial variance. Due to exponential expansion of the germline from the small number of founding cells, modest variation of cell cycle parameters between individual founding cells would be amplified into substantially higher levels of variance at later stages of development. Using a stochastic model to reconstruct the germ-cell development in human females, we show that the large variance observed in counts of human oocytes is consistent with the initial origin of these cells from a much smaller FGC population than is often assumed.

2. Overview
Our approach enables us to derive an estimate of the initial FGC population size on the basis of reliable data describing the number of oocytes present at various later stages of human development, when ovarian tissue is more readily available for analysis. Cytological counts of oocytes have shown not only that the size of the female germ cell population varies substantially between individuals of the same age, but also that its age-dependent magnitude is biphasic [13]. Numbers of germ cells increase during the first half of fetal development and then begin a progressive decline that extends throughout the reproductive years. The initial phase initiates with the separation of the germline from the soma, probably taking place no later than the peri-implantation stage (about 9 days after fertilization) [13,14]. Following this initial establishment, mitotically active oogonia undergo an exponential increase in number while only a small fraction of them show any sign of degeneration. At about five months of fetal development (5/12 = 0.42 year), the population reaches its peak as the oogonia enter into meiotic arrest. The germ cells (now defined as primary oocytes) become invested by layers of nurturing granulosa cells to form the follicles, which are readily recognized and enumerated by microscopic examination. The second phase of female germ line development, spanning the period from the late embryonic stage to the onset of menopause in adult females (t = 0.42 – 52 years), is characterized by a progressive decline in the number of follicles, largely due to apoptosis [15]. This decline is approximately exponential, but is accelerated in women older than age 38 [16]. Among the million or so oocytes present late in fetal development of the mother, the vast majority will undergo apoptosis while only 300–400 will progress fully through maturation and undergo ovulation during the woman’s reproductive life. Eventually, when the number of oocytes in the resting pool falls below 1000, menopause occurs [16].

3. Germ-Cell Kinetics
Oogonium-Birth Model
The stochastic model we use to describe germline development consists of two separate dynamic components (Figure 1). During the early embryonic stage (t = 0 – 0.42 year), a pure-birth model [11] can be used to describe the rapid proliferation of oogonia. At time t = 0, the germline is founded by a ancestral cells (FGCs) that are newly separated from the soma. At time t (0 ≤ t ≤ 0.42), the number X(t) of oogonia follows a negative-binomial probability distribution [11]:

$$\Pr[X(t) = n] = \left(\begin{array}{c} n-1 \\ a-1 \end{array}\right) \left(e^{-\lambda t}\right)^a \left(1 - e^{-\lambda t}\right)^{n-a}. \quad (1)$$

Figure 1
The number of female germ cells in humans undergoes three distinct rate changes, as diagrammed here and defined in the model. For the sake of clarity, the age coordinate is expanded artificially during the embryonic phase. The proliferative phase initiates at the time of germline-soma separation (ca. 9 days after fertilization; t = 0.0 year) and ends after 5 months of gestation (t = 0.42). The declining phases begin later in fetal life and continue into adulthood (t = 0.42 – 50 years) with an accelerated rate of oocyte depletion beginning at age 38 [16]. The dotted line shown during the embryonic stage emphasizes that oogonial cell counts from this period are inaccessible to reliable determination.
Here $\lambda$ is the oogonial division rate. The expectation and variance of the random variable $X(t)$ are $E[X(t)] = ae^{\lambda t}$, $\text{Var}[X(t)] = ae^{2\lambda t}(e^{\lambda t} - 1) = ae^{2\lambda t}$, respectively. Note that the approximation holds true if $ae^{\lambda t}$ is large (for humans, $ae^{\lambda t} > 10^4$). The moment ratio $a = E^2[X(t)]/\text{Var}[X(t)]$ then yields an estimate of the number of FGCs.

If one were dependent on using the pure-birth model to estimate $a$, this could be accomplished by collecting gonadal tissues from a series of abortices and establishing the total number of oogonia in each specimen ($x_i(t)$, $i = 1, 2, ..., I$) by microscopic evaluation. However, reliance on access to fetal tissue clearly has several drawbacks. First, many spontaneous abortions are associated with chromosomal aberrations [17] and may therefore display an abnormal pattern of growth kinetics. Second, access to non-diseased fetuses for research is limited by ethical concerns. And, third, microscopic examination of fetal tissues from an early stage of pregnancy is technically challenging. The boundaries of fetal gonad are not clearly demarcated from surrounding cell types and the oogonial cells are difficult to distinguish from the somatic cells. On the other hand, the ovarian follicles that arise at later stages of development are cytologically distinct and can be enumerated with precision. The following derivation of a pure-death model for germ cell dynamics enables us to use this more precise enumeration to advantage.

**Oocyte-Death Model**

A pure-death model, as described by Bailey [11], can be used to obtain an explicit formulation for the declining phase of germ cell numbers after proliferation has ceased and the apoptotic decline has begun ($t = 0.42 - 52$ years). Consistent with the findings of Faddy et al. [16], we permit the rate of oocyte loss from the resting pool to vary with age $t$. The cumulative rate function $f(t)$ is defined as follows:

$$f(t) = \begin{cases} 
\mu_1(t - 0.42), & 0.42 \leq t < 38 \\
\mu_1(38 - 0.42) + \mu_2(t - 38), & 38 \leq t < 52
\end{cases}$$

(2)

Conditional on the initial number $X(t) = n$ of oocytes at $t = 0.42$ year, the number of oocytes in the resting pool at age $t$ ($0.42 \leq t \leq 52$ years) now follows a binomial probability distribution [11]:

$$P[X(t)] = \begin{pmatrix} a \\ n \end{pmatrix} e^{-\lambda t} \sum_{i=1}^{n-1} \frac{(e^{\lambda t} - 1)}{(i-1)!} e^{-\mu_1 i}.$$

(3)

4. Estimation Methods

The unknown parameters ($a, \lambda, \mu_1, \mu_2$) can be estimated using two different methods. The moment-based method estimates only the number of FGCs (parameter $a$), while the maximum likelihood method estimates all four parameters ($a, \lambda, \mu_1, \mu_2$) simultaneously.

**Moment-Based Method**

As mentioned above regarding the oogonium-birth model, the random variable $X(t)$ follows a negative binomial distribution, and therefore the moment ratio $E^2[X(t)]/\text{Var}[X(t)]$ yields an estimate of $a$. This relationship holds true with oocyte depletion (oocyte-death model) following the period of exponential growth. As a verification, notice first that the probability-generating function for the negative binomial probability distribution at $t = 0.42$ is $P_{X(t)}(s) = \{1 - e^{-\lambda t} (1 - s)\}^{-a}$. Therefore, the probability-generating function for the binomial probability distribution at $t > 0.42$ (conditional on $X = n$) is $P_{X(t)}(s) = (1 - e^{-\lambda t} + e^{-\lambda t} s)^n$. The compounded probability-generating function ($t > 0.42$) is then given by $P_{X(t)}(s) = \left\{1 + \frac{e^{-\lambda t} - f(t)}{1 - e^{-\lambda t}} (1 - s)\right\}^{-a}$. The mean and variance for oocyte population size in adult women can be derived from the first and second derivatives of $P_{X(t)}(s)$:

$$P_{X(t)}(s) = E[X(t)] = \lambda \cdot e^{-\mu_1 t},$$

$$\text{Var}[X(t)] = \lambda \cdot e^{-\mu_1 t} \cdot (1 - e^{-\mu_1 t}).$$

Maximum Likelihood Estimation (MLE)

Although the moment ratio is simple to compute, its derivation requires a large sample size and provides no estimates of the oogonium-birth rate ($\lambda$) and the oocyte-death rates ($\mu_1, \mu_2$). The maximum likelihood method is not subject to these difficulties. To derive the likelihood function, note first that Equation 3 (pure-death model) is a probability function conditional on $P[X(t)] = n$ (Equation 1; pure-birth model). For each of the oocyte counts obtained from a series of autopsies ($x_i(t)$, $i = 1, 2, ..., I$) during the post-embryonic stage ($0.42 \leq t \leq 52$ years), we can combine them to define:

$$L(a, \lambda, \mu_1, \mu_2) = \prod_{i=1}^{I} \left\{ \begin{pmatrix} a \\ x_i(t) \end{pmatrix} e^{-\lambda t} \cdot \sum_{i=1}^{n-1} \frac{(e^{\lambda t} - 1)}{(i-1)!} e^{-\mu_1 i} \right\}.$$

(4)

Since each observation $x_i(t)$ makes a contribution like $L(x_i \mid a, \lambda, \mu_1, \mu_2)$ to the likelihood, the final likelihood $L$ of the entire data set is the product of all such terms ($0.42 \leq t_i \leq 52$ years), such that $L = \prod L(x_i(t) \mid a, \lambda, \mu_1, \mu_2)$. To evaluate the likelihood in Equation 4 numerically. In practice, we first generate samples of $n$ from this distribution and then sum the values of the negative binomial probabilities for $a$ given the sampled values of $n$. We obtain stable likelihood estimates with as few as 100 samples of $n$. During the process of searching the parameter space, we also restrict the parameter $a$ to be a positive integer (i.e., $a = 1, 2, 3, ...$).
5. Analysis of Oocyte-Counting Data

Using the above algorithms, we have analyzed published oocyte counts from 102 human females [16,19]. These computations yield MLEs of the four parameters, which are: \( \hat{a} = 2 \), \( \hat{\lambda} = 31.2/\text{year} \) (95% CI: 30.3 – 31.9), \( \hat{\mu}_1 = 0.079/\text{year} \) and \( \hat{\mu}_2 = 0.248/\text{year} \). The 95% confidence intervals (95% CI) are based on Markov chain Monte Carlo methods [20]. The expected number of oocytes (solid curve) and the pointwise 80% CI (shaded region) for the predicted counts on the basis of the model are shown together with oocyte counts in Figure 2.

The large MLE of \( \hat{\lambda} \) justifies our use of the moment-based method to estimate the parameter \( a \). To enable moment-based analysis we have grouped the data into 5 discrete age intervals [16] and calculated the mean and variance for each age interval. Computing the moment ratio index for each age interval yields the values shown (Table 1) and an overall average of \( \hat{a} = 2.7 \). This value agrees well with the MLE and with those values (\( \hat{a} = 2 – 9 \)) that were derived from the segregation of genetic markers in the mouse [4-10].

Table 1: Oocyte-counting data from Faddy et al. [16].

| Age       | Sample Size | Mean | S.E. | Ratio Index \( \hat{a} \) |
|-----------|-------------|------|------|---------------------------|
| 19–30 years | 5           | 78980 | 15580 | 5.1                       |
| 31–35 years | 13          | 25300 | 4860  | 2.1                       |
| 36–40 years | 14          | 21450 | 2650  | 4.7                       |
| 41–45 years | 32          | 7320  | 1450  | 0.8                       |
| >45 years   | 36          | 1880  | 310   | 1.0                       |

6. Discussion

Our estimated values of the founding germ cell number \( a \) (MLE \( \hat{a} = 2 \), moment estimate \( \hat{a} = 2.7 \)) differ substantially from the estimates made in previous studies [2,3] that relied on cytochemical staining of embryonic material. This discrepancy might best be explained by assuming that not all cells sharing the same cytochemical phenotype (alkaline phosphatase staining) in common with a FGC population proceed through that course of development. This view is supported by the observation that alkaline phosphatase is present not only in germ cells of the mouse, but also in somatic cells that surround these germ cells [21]. Recent description of cytochemical markers that are more specifically restricted to the germ cells confirms that the expression of alkaline phosphatase occurs in a wider spectrum of cell types [22].

Genetic analysis provides a more stringent approach to germ cell enumeration in the mouse, where experimental crosses permit reliable determination of marker transmission. If a mutation or other stably transmissible cellular property arises early in development, only a subset of cells will display the mutation at later stages. Such mosaic presence of the trait provides an opportunity to help define the stage at which the germ cell precursors (FGCs) had become segregated from the predominant somatic cell population [23,24]. If the extent of mosaicism were closely similar between soma and germline, this would indicate that the size of the FGC population is large, since it would have provided a representative sample of the mosaicism originally present throughout the early embryo. To the contrary, if the correlation between somatic and germline mosaicism is weak, the number of FGCs must be limited and the transmission of any somatic markers to offspring should be more stochastic. This type of analysis [4-10] generally predicts small number of FGCs, consistent with our derivation from the stochastic modeling of oocyte counts.

The present statistical analysis indicates that germ cell proliferation occurring during embryogenesis is characterized both by a small FGC population (\( \hat{a} = 2 \)) and by a rapid...
rate of cell division (\( \hat{\lambda} = 31.2/\text{year} \)). Although we have estimated all four parameters of the model (\( a, \lambda, \mu_1, \) and \( \mu_2 \)) in concert using the oocyte-counting data collected post-embryonically, the division rate \( \lambda \) can be verified independently by microscopic study of embryonic tissues. In a recent publication, Bendsen et al. [25] reported their evaluation of 10 fetal gonad specimens between the ages of 6 and 9 weeks. Using morphological clues to distinguish germ cells from somatic cells, these authors were able to establish the number of germ cells per tissue sample throughout this period. We used exponential regression against fetal age \( t \) to analyze the data of Bendsen et al. and obtained a direct estimate of \( \hat{\lambda} = 35.4/\text{year} \) (95% CI: 17.2/year – 53.7/year). Thus, our estimate of the division rate of the FGC is in good agreement with the experimental data.

It has recently been reported [26] that mammalian ovaries contain stem germ cells that are competent to enter into mitosis at a late stage of development. If confirmed by further work, this finding would necessitate some modification of the model proposed here. However, additional computation suggests that the potential effect on the estimation of the model proposed here. This coincident cause of this variance.

Finally, we note that the squared root of an inversed moment ratio is mathematically equivalent to the coefficient of variation – the quotient of the standard deviation divided by the mean. This parameter is commonly used in biostatistics to characterize the extent of random variation across a broad range of biological processes. This coincidence suggests that the founder-effect interpretation that we have proposed may have broader applications.

**Competing interests**
The author(s) declare that they have no competing interests.

**Authors’ contributions**
CJZ carried out the initial mathematical derivations and analyzed the oocytes-counting data using the moment-ratio method. EGL and SHM extended the mathematical derivations and completed the maximum-likelihood estimation. BB reviewed the biological implications of the models. All authors participated in preparations and revisions of the manuscript.

**Acknowledgements**
We would also like to acknowledge financial support from the National Institutes of Health (BB, EGL & SHM) and the National Institute of Occupational Safety and Health (CJZ).

**References**

1. Hubbard EJA, Pera RAR: A germ-cell odyssey: fate, survival, migration, stem cells and differentiation. Meeting on germ cells. EMBO Rep 2003, 4:352-7.
2. Lawson KA, Hage WJ: Clonal analysis of the origin of primordial germ cells in the mouse. Ciba Foundation Symposium 1994, 182:68-91.
3. McMahon A, Fosten M, Monk M: X-chromosome inactivation mosaicism in the three germ layers and the germ line of the mouse embryo. J Embryol Exp Morphal 1983, 74:207-20.
4. Mintz B: Hermaphroditism, sex chromosomal mosaicism and germ cell selection in allophenic mice. J Animal Sci 1968, 27(suppl 1):51-60.
5. Mintz B: Gene control of mammalian differentiation. Annu Rev Genet 1974, 8:411-70.
6. Russell LB: Genetic and functional mosaicism in the mouse. In Role of Chromosomes in Development Edited by: Locke M. Academic, New York; 1964:153-81.
7. Russell LB: Analysis of the albino-locus region of the mouse. II. Mosaic mutants. Genetics 1979, 91:141-7.
8. Searle AG: Evidence from mutable genes concerning the origin of the germline. Basic Life Sci 1978, 12:209-24.
9. Soriano P, Jaenisch R: Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell linkages. Cell 1986, 46:19-29.
10. Wilkie TM, Brinster RL, Palmiter RD: Germline and somatic mosaicism in transgenic mice. Dev Biol 1986, 118:9-18.
11. Bailey NTJ: The elements of stochastic processes with applications to the natural sciences John Wiley & Sons, New York; 1964.
12. Yule GU: A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Proc R Soc London Ser B 1924, 213:21-87.
13. Gosden RG: Ovulation I: Oocyte development throughout life. In Gametes – The Oocyte Edited by: Grudzinskas JG, Yovich JL. Cambridge University Press, Cambridge; 1995:119-149.
14. Wilcox AJ, Baird DD, Weinberg CR: Time of implantation of the conceptus and loss of pregnancy. N Engl J Med 1999, 340:1796-9.
15. Tilly JL: Commuting the death sentence: how oocytes strive to survive. Nat Rev Mol Cell Biol 2001, 2:838-48.
16. Faddy MJ, Gosden RG, Gougeon A, Richardson SJ, Nelson JF: Accelerated disappearance of ovarian follicles in mid-life: implications for forecasting menopause. Hum Reprod 1992, 7:1342-6.
17. Boue J, Bou A, Lazar P: Retrospective and prospective epidemiological studies of 1500 karyotyped spontaneous human abortions. Teratology 1975, 12:11-26.
18. Morgan BJ: Applied Stochastic Modelling Arnold Texts in Statistics, Oxford University Press; 2000.
19. Leidy LE, Godfrey LR, Sutherland MR: Is follicular atresia biphasic? Fertil Steril 1998, 70:851-9.
20. Gilks WR, Richardson S, Spiegelhalter DJ, eds: Markov Chain Monte Carlo in Practice Chapman & Hall, New York; 1996.
21. MacGregor GR, Zambrowicz BP, Soriano P: Tissue non-specific alkaline phosphatase is expressed in both embryonic and extraembryonic lineages during mouse embryogenesis but is not required for migration of primordial germ cells. Development 1995, 121:1487-96.
22. Saitou M, Barton SC, Surani MA: A molecular programme for the specification of germ cell fate in mice. Nature 2002, 418(6895):293-300.
23. Nesbitt MN, Gartler SM: The Applications of Genetic Mosaicism to Developmental Problems. Annu Rev Genet 1971, 5:143-162.
24. McLaren A: Numerology of development. Nature 1972, 239(5370):274-6.
25. Bendsen E, Byskov AG, Laursen SB, Larsen HP, Andersen CY, Westergaard LG: Number of germ cells and somatic cells in human fetal testes during the first weeks after sex differentiation. Hum Reprod 2003, 18:13-8.
26. Johnson J, Canning J, Kaneko T, Pru JK, Tilly JL: Germline stem cells and follicular renewal in the postnatal mammalian ovary. Nature 2004, 428:445-50.