GENETICS

The contribution of genetic and environmental factors to the duration of pregnancy

Timothy P. York, PhD; Lindon J. Eaves, PhD; Michael C. Neale, PhD; Jerome F. Strauss III, MD, PhD

This review describes how improvements in biometric-genetic studies of twin kinships, half-sibships, and cousinships have now demonstrated a sizeable fetal genetic and maternal genetic contribution to the spontaneous onset of labor. This is an important development because previous literature for the most part reports only an influence of the maternal genome. Current estimates of the percent of variation that is attributable to fetal genetic factors range from 11—35%; the range for the maternal genetic contribution is 13—20%. These same studies demonstrate an even larger influence of environmental sources over and above the influence of genetic sources and previously identified environmental risk factors. With these estimates in hand, a major goal for research on pregnancy duration is to identify specific allelic variation and environmental risk to account for this estimated genetic and environmental variation. A review of the current literature can serve as a guide for future research efforts.

Key words: duration of pregnancy, environment, gene, genetic epidemiology, gestational age at birth, preterm birth

"Duration of pregnancy and intrauterine growth are dark areas in human biology"[1]

Births at <37 completed weeks of gestation are preterm and account for most perinatal deaths and morbidity.[2,3] Preterm birth has been shown to be associated with adverse outcomes early in life through adulthood.[4-11] In humans, spontaneous preterm labor in infants without congenital abnormalities has been described as a culmination of a series of physiologic and anatomic changes in both the mother and fetus.[12-15] Initial evidence for the identity of these factors was derived from reports that preterm birth is correlated among successive births of the same mother and between other familial relationships.[1,16] The etiologic factors that explain these correlation patterns could be genetic factors that are shared among relatives, environmental factors that are shared within a family, or both.

Perhaps the most fascinating etiologic aspect of the duration of pregnancy is the consideration of both the fetal and maternal genomes, yet the interplay between the in utero environmental effect of the maternal genotype and the developmental effect of the fetal genotype presents challenges to the understanding of pathophysiologic mechanisms. On the other hand, environmental sources of variation can be partitioned into familial sources into those that influence all births of the same mother (eg, socioeconomic status) and those that are unique to individual pregnancies (eg, infectious disease). Biometric genetic theory and conceptual models that have been developed over the past 100 years provide the framework to test competing hypotheses regarding genetic and environmental contributions to the duration of pregnancy. Considering the magnitude of the public health impact of preterm birth, surprisingly few of these studies have been conducted to quantify these sources. Although these studies have provided compelling evidence for genetic influence on individual differences in the duration of pregnancy, many basic questions remain as to the nature of these sources that are fundamental to continued research efforts. In this review, we assess current genetic epidemiologic findings in this area to address the following questions regarding the spontaneous onset of labor: (1) does the fetal and/or maternal genome contribute to differences in the duration of pregnancy? (2) To what extent do environmental factors contribute differences in the duration of pregnancy, and are these more/less important than genetic factors? (3) Is there heterogeneity in the effect of these factors across self-identified racial/ethnic groups that could account for known differences in preterm birth rates?

Genetic epidemiologic research and the duration of pregnancy

Birth outcomes research provides a fascinating model of inquiry, potentially involving direct effects of and interaction between the fetal and maternal genomes. Yet, before specific factors underlying the duration of pregnancy can be examined, standard practice in genetic epidemiologic research is to use a top-down gene finding approach to first estimate the overall extent to which genetic factors influence the phenotype of interest and whether these are of fetal origin, maternal origin, or both (Appendix).

From the Virginia Institute for Psychiatric and Behavioral Genetics (Drs York, Eaves, and Neale) and the Departments of Human and Molecular Genetics (Drs York, Eaves, Neale, and Strauss), Obstetrics and Gynecology (Drs York and Strauss), and Psychiatry (Drs Eaves and Neale), Virginia Commonwealth University School of Medicine, Richmond, VA.

Received Aug. 5, 2013; revised Sept. 26, 2013; accepted Oct. 1, 2013.

tcgqand HD073555 National Institutes of Health grants P60MD002256 (T.P.Y, J.F.S), R01HD034612 (J.F.S.)

The authors report no conflict of interest.

Reprints: Timothy P. York, PhD, Virginia Institute for Psychiatric and Behavioral Genetics, Department of Human and Molecular Genetics & OB/GYN, Virginia Commonwealth University, Box 980003, Richmond, VA 23298-0003. tpyork@vcu.edu.

0002-9378/$36.00
Published by Mosby, Inc.
http://dx.doi.org/10.1016/j.ajog.2013.10.001

The classic twin design has been applied widely to estimate the contribution of genetic and environmental factors that range from anthropomorphic traits to psychiatric outcomes.[17] For studies of birth outcomes, these models have included extended family structures to avoid the biases inherent in twin births that restrict generalizability to singleton births. Including the offspring of twins, siblings, and half-siblings allows for a wider array of hypotheses to be tested and has been shown to increase statistical power.[18] The information necessary to separate genetic from environmental effects is contained in the covariances of relationships that vary in their biologic relatedness,[19] As an example, cousins who are related through brothers on average share one-eighth of their genes and therefore a rough estimate of fetal genetic influences (because the mothers of these cousins are unrelated) can be derived by calculating the correlation between these collateral relatives. Although inspection of familial correlations provides a useful summary of the magnitude of genetic and environmental influence, these parameters typically are estimated across several biologic relationships simultaneously with the use of structural equation modeling techniques.[20]

Throughout this review, we adopt the convention that preterm birth is a clinically meaningful threshold that is imposed on the continuum of gestational age. Other informative thresholds have been proposed, including very early preterm birth (births at <32 weeks' gestation) and postterm birth (births at >42 weeks' gestation). Unless otherwise demonstrated, the choice of threshold does not necessarily imply anything fundamental about the underlying genetic mechanisms; it merely reflects a current judgment about when medical intervention is likely to be beneficial and cost-effective. Decisions about "where to draw the line" change as medical research obtains better information about the long-term consequences of particular trait levels. As understanding advances so do the standards of assessment and the criteria for clinical intervention.

Genetic contributions to the duration of pregnancy

Genetic variance is a function of allele frequencies and their effect sizes, both of which can vary across populations. Similarly, the influence of environmental factors may differ between populations. Therefore, it should be cautioned that estimates of heritability from one population do not necessarily predict heritability in other populations. Because heritability is a ratio of variances (Appendix) and although the amount of genetic variance could be constant across populations, the total variance could differ because of differences in the environmental variances and result in population-specific heritability estimates. For example, human height has been shown to be highly heritable, with estimates that range from 87–93% across several Western countries.[21] In poorer countries, the proportion attributable to environmental sources is larger and results in lower heritabilities.[22]

Heritability estimates for the duration of pregnancy in European and European American samples support the influence of both fetal and maternal genomes (Table). The contribution of fetal genetic factors range from 11–35%; the range for the maternal genetic contribution is 13–20%. The Norwegian[23] and Swedish[19] studies listed are consistent in their genetic proportions; the US sample[24] of individuals self-described as European American report a much larger proportion of fetal genetic variance than the European samples. The interpretation of genetic proportions across studies is limited because the total variance of gestational age can differ between samples and would require the unstandardized parameter estimates to be reported on the same scale for each study. Nevertheless, these studies importantly demonstrate (1) the consistency of both fetal and maternal genetic effects in explaining interindividual differences in gestational age at birth and (2) that these estimates are as large as or larger than environmental influences shared among births.

The 2 remaining studies that are listed in the Table partition the variance of gestational age at birth by dichotomizing the data at clinically meaningful thresholds. The first study classifies births as preterm if it occurs at <37 completed weeks' gestation,[25] and the second study classifies births that have occurred at >41 completed weeks' gestation (postterm).[26] The results from these 2 studies differ considerably in the proportion of fetal genetic and familial environmental influence, which suggests that these factors may vary in their effects across the range of gestational age. Although the attempt here is to diagnose different contributions at particular clinical definitions, it is not clear that moving the threshold provides additional information about genetic and environmental mechanisms under the assumption of a single underlying continuum of risk. This approach generally would lack the modeling framework to test whether variance components differ significantly at either end of the gestational age distribution. We have shown empirically and by simulation studies that imposing thresholds in the tails of the gestational age distribution results in genetic and environmental parameter estimates that vary widely because of a marked decrease in their precision, especially for more extreme thresholds.[19]

In epidemiologic studies, there is a high price associated with imposing a threshold on an otherwise continuous phenotype such as gestational age at birth, because it provides no gain in information and usually results in dramatically lower statistical power.[27,28] Dichotomization classifies distinct groups of individuals on the phenotype of interest and argues that there exists a natural boundary between these groups. Although conceptually attractive, this position can be difficult to justify empirically, because it eliminates variation within groups. An alternative viewpoint suggests that individuals, rather than belonging to discrete groups, could be a "mixture" of groups. For instance, it is possible that the overall distribution of gestational age at birth is comprised of 2 underlying risk distributions, each describing a different model of genetic risk. This might involve an increase in genetic risk with increasing gestational age or a multiplicative model of risk. The point of this discussion is not to support 1 mechanism or the other, but