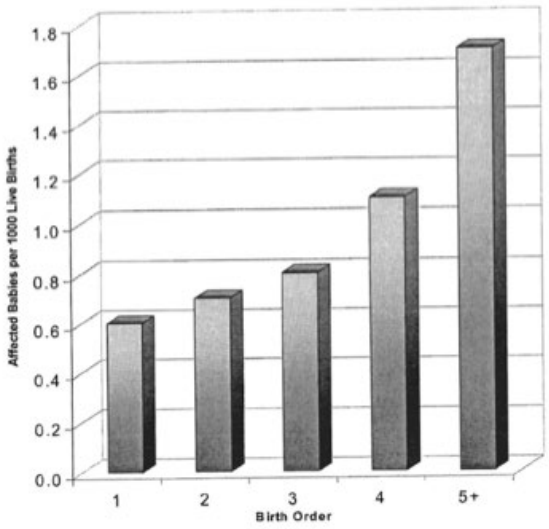
SISG 2021: Module 10
Session 2: Introduction to Genetic Epidemiology

The following questions are based on this figure:

Confounding example: Birth order and Down syndrome



Data from Stark and Mantel (1966) Source: Rothman 2002

1. **Can you think of a factor that would confound the observed association?**
2. **How can you use data on your proposed confounding factor to reassess the association between birth order and Down Syndrome?**
3. **Can you think of potential confounders in genetic epidemiology?**

1. One of the most significant factors confounding the observed association is the maternal age. It is associated with both the exposure (Birth order) and the outcome (Incidence of Down Syndrome)
2. One approach is to stratify the study population based on the maternal age and investigate the association of interest within each stratum.
3. When studying the genotype-phenotype association in genetic epidemiology research, the most significant confounder is population structure, which represents the difference in allele frequency between the subpopulations.

SISG 2021: Module 10
Session3: Human Genetic Variation

1. A recent study sequenced the genome of 2,504 individuals and identified 84.7 million SNPs (single nucleotide polymorphisms) between the participants. On average, each individual carried 3.5-4.3 million SNPs each. About 0.5% of those SNPs were in coding regions of genes. Remember, 1.5% of the genome is in a coding region. Why might only 0.5% of variants be in coding regions compared to what would be expected if SNPs were randomly allocated throughout the genome?

Heterogeneity of SNPs distribution across the genome suggests the coding regions are more conservative. The polymorphisms in coding region may lead to severe functional consequences (missense mutations, nonsense mutations, frameshift mutations), and may be eliminated before they get the chance to be passed to the next generation.

2. Match the genetic term with the definition:

a. Nonsense
b. Heterozygous
c. Exon
d. Allele
e. Synonymous
f. Missense
g. Non-coding region
h. Haplotype
i. Autosomal
j. Phenotype
k. Genotype
l. Frameshift
m. Intron
n. Homozygous

d___ Alternative forms of a gene or DNA base.
k___ Genetic makeup of an individual at a particular DNA location based on both alleles.
b___ Genotype consisting of two different alleles at a particular location.
e___ DNA base change that does not change the translated amino acid.
n___ Genotype consisting of two of the same alleles at a particular location.
j___ Observable characteristics resulting from a genotype.
i___ Concerning the 22 pairs of chromosomes that are not sex chromosomes.
m___ Portion of gene that does not code for amino acids and appears in between exons.
l___ Insertion or deletion mutation that changes the whole subsequent sequence of amino acids by changing the 3-codon groups for generating amino acids.
c___ Portion of gene that encodes amino acids.
g___ Section of DNA that does not become protein.
a___ Substitution of a single DNA base that causes a stop in protein production.
f___ DNA base change that changes the translated amino acid.
h___ Set of DNA variations at several positions that are inherited together.

3. Look up "rs7412" in dbSNP (https://www.ncbi.nlm.nih.gov/snp/).
   https://www.ncbi.nlm.nih.gov/snp/rs7412
   a. What are the DNA bases identified at this location? C/T
   b. What gene is this SNP located in? *APOE*
   c. What is the effect of this SNP on the amino acid sequence? Missense Mutation
   d. Click on the "frequency" tab to the left. What is the frequency of the minor allele (less common allele) in the 1000 Genomes study overall? How do these frequencies differ by ancestral subgroup within this study? Total: 92.5% C/7.5% T