MOLECULAR EPIDEMIOLOGY

Afiono Agung Prasetyo
Faculty of Medicine Sebelas Maret University Indonesia

GENERAL EPIDEMIOLOGY

General epidemiology is the scientific basis of public health

Descriptive epidemiology: distribution of disease in populations

- Incidence and prevalence rates
- Characterize the burden of disease in a population

Analytical epidemiology: determinants of disease

- Relative risks and odds ratios
- Characterize the risk factors for disease in populations

MOLECULAR EPIDEMIOLOGY

Based on general epidemiology

- Utilizes the same designs (cross-sectional, case-control, and prospective cohort)
- Utilizes molecular biology to define the distribution of disease in a population (descriptive epidemiology) and identify its potential etiologic determinants (analytical epidemiology)

Molecular epidemiology enhances our understanding of the pathogenesis of disease by identifying specific pathways, molecules and genes that influence the risk of developing disease; improves validity and reduces bias for assessments of environmental exposure; allows for evaluation of subclinical or early disease markers; reduces heterogeneity in the classification of diseases in descriptive studies; provides new standards for descriptive epidemiology; improves precision in analytical epidemiology.

Genetic / genomic / epigenomic epidemiology

Susceptibility genes, infectious agents and environmental risk factors

With molecular epidemiology, it is possible to evaluate a spectrum of molecular markers--from susceptibility genes, to biological markers of exposure, to markers of preclinical disease to diagnostic markers.

Answering two types of questions:

- Contributing to the identification of cause–effect relationships between exposure to putative risk/preventive factors and disease
- Measuring the exposure–disease association (strength, dose–response, population impact).

DESCRIPTIVE + ANALYTICAL MOLECULAR EPIDEMIOLOGY

- estimate individuals' relative risk of disease (magnitude of the increase in risk associated with the presence of a specific molecular marker)
- Determine their absolute risk of disease (probability of developing the disease if the molecular markers under evaluation are present)
  - Essential for the development of medical and public health strategies for disease prevention
  - Require knowledge of the general epidemiology of the disease in the population, particularly incidence/prevalence rates
- Evaluation of the population attributable risk (amount of disease in a population that could be prevented by intervening on the specific molecular marker under evaluation)

- Essential for preventive medicine and public health approaches

- Information of relative, absolute and attributable risk of disease, associated with the presence of specific molecular markers
  - Necessary to develop molecular diagnostic or disease prevention strategies

CONSIDER PRIOR TO INITIATING MOLECULAR EPIDEMIOLOGIC STUDIES

- Is available, valid, reproducible, and can be standardized?
- Is performed on biological specimens that are accessible?
- Is socially and ethically acceptable?
- Is rapid and cost-effective?
- Is highly sensitive and specific?

GENETIC EPIDEMIOLOGY

Afiono Agung Prasetyo
Faculty of Medicine Sebelas Maret University Indonesia

HOW AND WHY SOME HEALTH OUTCOMES CLUSTER IN FAMILIES?

- Does the trait or disease under study cluster in families?
- Are there combinations of diseases that cluster in families?
- If so, what is the relative influence of genetic and environmental factors?
- What are the specific genetic variants and environmental factors influencing the trait or disease?
- How do the environmental and genetic factors interact?

Notes…

- The same disease can have a broad clinical spectrum, and the same clinical features can have very different etiologies.
- It is critical that there are well-defined criteria for the disease or trait under study.
- If possible, genotyping should be done without knowledge of the disease and phenotyping without information on genotypes (blinding) because both the genotyping and the phenotyping come with some measurement errors.

CLUSTERING OF TRAITS AND DISEASES IN FAMILIES

For diseases with manifestation late in life:

- Many family members will be too young to be at risk of having the disease
- Others die before the typical time of diagnosis

In studies where the participant needs to return a questionnaire and participate in interviews about family occurrence:

- There is a risk that families with several affected members are more likely to participate or be detected because these families may be particularly interested in studies

IS THERE AN INCREASED RISK OF DISEASE IN RELATIVES OF CASES?

In general, more distant relatives will have lower risks than close relatives, since closer relatives are more likely to share genetic risk factors with the case (‘proband’) than are more distant relatives.

Restricted to first degree relatives (parents, siblings and offspring).

IS THE FAMILIAL AGGREGATION DUE TO GENES OR ENVIRONMENT?

A. Twin study: comparing the effect of having different genetic make-up (as for DZ pairs) with that of having the same genetic makeup (as for MZ pairs) while both twins of a pair are brought up in the same household

In humans two types of twinning:

- monozygotic: share all their genetic material
- dizygotic : average share 50%

Does not identify specific genes that affect the trait but rather assesses the overall effect of genetic factors: the degree to which differences in the phenotype are attributable to genetic differences between people.

Alzheimer’s disease (not early onset):

- monozygotic twin has a 60–80% risk of becoming affected, while the risk is 30–40% if the pair is dizygotic.

Parkinson's disease: 5% for both monozygotic and dizygotic twins.

B. Adoption studies: comparing the frequency of disease among the offspring of unaffected parent whose children have been adopted out with the frequency in children who have unaffected biological parents but an affected adoptive parent

Adoptees share genetic variants with their biological parents but not the parents' environment, and they share the environment to some extent, but not gene variants, with their adoptive families.

Heston's 1966 study: 47 children who had schizophrenic mothers and who were put up for adoption:

- 5 developed schizophrenia
- none of the 50 control adoptees developed schizophrenia
- schizophrenia has a strong genetic component

C. HALF-SIB STUDIES

In countries with population registers

It is possible on a nationwide level to identify individuals who have changed their spouse or residence (or other environmental factors)

Information from these registers can be used to set up a study that is particularly well suited for studying nature–nurture effects on reproductive outcomes or diseases in early childhood.

BENEFIT OF FAMILY-BASED ASSOCIATION STUDIES

The control for confounding bias due to population stratification

Family members may be easier to recruit for some disorders than unrelated individuals

The potential for quality assurance measures when the same data are collected on more than one family member

If genotyping is performed, quality assurance measures such as checking for Mendelian inheritance are also possible.

EXPOSURE–DISEASE ASSOCIATIONS THROUGH STUDIES OF RELATIVES

Nearly all lung cancers occur in smokers.

- It is causal or a result of confounding, e.g., genetic?

Twin studies (i.e., twin pairs where one twin is a smoker and the other is not)

If the smoking–lung cancer association is caused by genetic factors:

- Among monozygotic twins: the smoking co-twin to get lung cancer is not more often than the nonsmoking co-twin.
- For dizygotic twins: a weaker association than observed in the general population because we partly control for genetic factors (dizygotic twins share, like siblings, about 50% of their genes)

If is not due to genetic or other familial factors: the smoking co-twins have the highest lung cancer risk regardless of zygosity

GENE–ENVIRONMENT INTERACTION

G6PD deficiency

- An X-linked trait of the enzyme glucose-6-phosphate dehydrogenase deficiency that facilitates energy metabolism in cells (like red blood cells) and helps protect the cell from oxidative damage
- Can result in hemolysis and anemia if the affected are exposed to certain nutritional insults such as Fava beans or pharmacologic agents including some antibiotics and antimalarias.
- Thus individuals with a risk genotype are normal in the absence of an environmental exposure

SOURCES OF ERROR IN GENETIC EPIDEMIOLOGY

Multiple Testing

- False positive findings
  Tests of a million SNPs would be expected to generate 50,000 SNPs showing $P < 0.05$ by chance alone
- New analytic approaches (the false discovery rate) can assist in minimizing to a reasonable number the signals that require additional investigation
- Not only the exposure side (genotypes) has contributed to the multiple testing problem, but also the outcome side (health, diseases, behaviors)

Population Stratification

Selection of the control group

- Gene variants that show a north–south Europe gradient (e.g., ApoE-4 or genes for lactose intolerance) will seem to be associated with height and skin color because there is also a north–south gradient in Europe in height and skin color
- Can be removed by matching on ethnic/ancestral background, restricting studies to highly homogeneous populations and by using family-based controls.
  Use the large numbers of markers (marker-based definitions of ancestry)

Laboratory Errors

- Laboratory testing is only as reliable as the samples accurately reflect their biological origins
- Mishandling and mislabeling
- Family histories (egg and embryo donation, adoption?
- DNA quality

Case and control DNA, or infant and parent DNA, may come from different biological sources

The behavior of DNA in genotyping assays is partly dependent on its origin and processing.

BIOSAMPLING METHODS

Afiono Agung Prasetyo
Faculty of Medicine Sebelas Maret University Indonesia

SAMPLE COLLECTION
Blood, urine, oral cells, sputum, breast milk, hair, toenails, saliva, meconium, feces, fat, frozen / paraffin tissues
Blood and urine are the most frequently collected because of the ease of collection and general acceptability

BLOOD SAMPLES
Timing of collection (time of day, month, season, fasting/nonfasting)?
Type of tube? Needles?
Anticoagulant? If so, of which type?
How much?
If it is to be shipped, under what conditions?

- Chill packs limit degradation of specific analytes but lead to lower viability of lymphocytes

Time to processing?
Sterile? Protected from sunlight? Specific additives?

URINE SAMPLES
First morning void samples are more concentrated while random, spot samples are easier to collect but may not be representative.
The most difficult to collect but also the most accurate is 24-hour collection.
Most frequently used for measurement of excreted chemicals or hormones.
Can be used for genotyping when blood or buccal cells are not available.

BUCCAL CELL SAMPLES
Swabs, brushes, or mouthwash solutions
It must be remembered that a significant portion of the DNA is from bacteria.

SAMPLE STORAGE
Blood or urine samples: -70C or -80C, or -196C
Liquid nitrogen freezers (< -130C) might be a better choice for very long-term storage of samples, such as in a prospective cohort study
Should be split between two freezers

SAMPLE TRACKING
documentation of sample characteristics:

- when sample was collected?
- sample was hemolyzed?
- sample volume?
- the samples in the storage system?
- freeze-thaw history?

any new subaliquots?
any samples were sent for analysis, and the assays to be conducted?

BIOMARKERS AS ENDPOINTS IN INTERVENTION STUDIES

Afiono Agung Prasetyo
Faculty of Medicine Sebelas Maret University Indonesia

BIOMARKER?

- 'a measurable, biological parameter that predicts the risk of human disease, disorders or conditions, but is not a measure of the disease, disorder or condition itself'
- to assess exposure to potential environmental hazards, to gain insight into disease mechanisms and to understand acquired or inherited susceptibility

IDENTIFICATION AND VALIDATION OF BIOMARKERS

If biomarkers are to provide an effective index of efficacy, they must provide a quantifiable index of reduction in hazard exposure, disease initiation or progression

To quantify the strength of the biomarker–disease association: using prospective cohort or case–control studies

USE OF BIOMARKERS IN MAKING HEALTH CLAIMS

- The completeness and appropriateness of the described methodology
- Appropriate and accurate description and quantification of exposure to the diet, food or food component
- Appropriate and accurate quantification of the health related outcome
- Sample size; Sample and measurement bias
- Potential confounding variables
- Inclusion of appropriate controls
- Study duration
- Appropriate statistical methods