Evaluating Potential Biases in and Interpreting Results from Epidemiologic Studies

U.S. Department of Health and Human Services
National Institutes of Health
National Human Genome Research Institute

Teri A. Manolio, M.D., Ph.D.
Director, Office of Population Genomics

Thomas A. Pearson, M.D., Ph.D.
Visiting Scientist

October 24, 2007

Nothing to Disclose
Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power 1 – $\beta$ (one minus the Type II error rate). The probability
Most Science Studies Appear to Be Tainted By Sloppy Analysis

We all make mistakes and, if you believe medical scholar John Ioannidis, scientists make more than their fair share. By his calculations, most published research findings are wrong.

Dr. Ioannidis is an epidemiologist who studies research methods at the University of Ioannina School of Medicine in Greece and Tufts University in Medford, Mass. In a series of influential analytical reports, he has documented how, in thousands of peer-reviewed research papers published every year, there may be so much less than meets the eye.
A Common Genetic Variant Is Associated with Adult and Childhood Obesity

Comment on “A Common Genetic Variant Is Associated with Adult and Childhood Obesity”

Ruth J. F. Loos, Inês Barroso, Stephen O’Rahilly, Nicholas J. Wareham

Herber et al. (Reports, 14 April 2006, p. 279) found that the rs7566605 genetic variant, located upstream of the INSIG2 gene, was consistently associated with increased body mass index. However, we found no evidence of association between rs7566605 and body mass index in two large ethnically homogeneous population-based cohorts. On the contrary, an opposite tendency was observed.

Association between minor allele of rs7566605 near INSIG2 and increased BMI and homozygosity in 923 related Framingham Heart Study (FHS) participants

Association reproduced in four additional cohorts

Not seen in fifth cohort
The Association of a SNP Upstream of INSIG2 with Body Mass Index is Reproduced in Several but Not All Cohorts

Helen N. Lyon1,2,3, Valur Emilsson4, Anke Hinney5, Iris M. Heid6,7, Jessica Lasky-Su8,9, Xiaofeng Zhu10, Gudmar Thorleifsson4, Steinunn Gunnarsdottir4, G. Bragi Walters4, Unnur Thorsteinsdottir4, Augustine Kong4, Jeffrey Gulcher4, Thuy Trang Nguyen11,12, André Scherag11,12, Arne Pfeufer13,14, Thomas Meitinger13,14, Günter Brönner5, Winfried Rief11,12, Manuel E. Soto-Quiros15, Lydiana Avila15, Barbara Klanderman8, Benjamin A. Raby8, Edwin K. Silverman8, Scott T. Weiss8, Nan Laird8, Xiao Ding8, Leif Groop16,17,18, Tiinamaija Tuom17,18,19, Bo Isomaa19, Kristina Bengtsson17,18, Johanna L. Butler1,2, Richard S. Cooper20, Caroline S. Fox21, Christopher J. O'Donnell21, Caren Vollmer6, Juan C. Celedón8, H. Erich Wichmann5,7, Johannes Hebebrand1, Kari Stefansson4, Christoph Lange4, Joel N. Hirschhorn2,22

A SNP upstream of the INSIG2 gene, rs7566605, was recently found to be associated with obesity as measured by body mass index (BMI) by Herbert and colleagues. The association between increased BMI and homozygosity for the minor allele was first observed in data from a genome-wide association scan of 86,604 SNPs in 923 related individuals from the Framingham Heart Study offspring cohort. The association was reproduced in four additional cohorts, but was not seen in a fifth cohort. To further assess the general reproducibility of this association, we genotyped rs7566605 in nine large cohorts from eight populations across multiple ethnicities (total n = 16,969). We tested this variant for association with BMI in each sample under a recessive model using family-based, population-based, and case-control designs. We observed a significant (p < 0.05) association in five cohorts but saw no association in three other cohorts. There was variability in the strength of association evidence across examination cycles in longitudinal data from unrelated individuals in the Framingham Heart Study Offspring cohort. A combined analysis revealed significant independent validation of this association in both unrelated (p = 0.046) and family-based (p = 0.004) samples. The estimated risk conferred by this allele is small, and could easily be masked by small sample size, population stratification, or other confounders. These validation studies suggest that the original association is less likely to be spurious, but the failure to observe an association in every data set suggests that the effect of SNP rs7566605 on BMI may be heterogeneous across population samples.
The Association of a SNP Upstream of \textit{INSIG2} with Body Mass Index is Reproduced in Several but Not All Cohorts

- Nine large cohorts from eight populations across multiple ethnicities
- Family-based, population-based, case-control designs
- Association at $p < 0.05$ in five cohorts but none in three cohorts
- Variability in strength of association over time
- Replication both in unrelated ($p = 0.046$) and family-based ($p = 0.004$) samples
- Suggests initial finding unlikely to be spurious but effect likely to be heterogeneous

### rs7566605 C/C Genotype and BMI > 30 kg/m² in Unrelated Individuals (Lyon et al, *PLoS Gen* 2007)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Obesity Association</th>
<th>Frequency C/C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Essen</td>
<td>1.75</td>
<td>[1.15-2.67]</td>
</tr>
<tr>
<td>FHS 1</td>
<td>1.26</td>
<td>[0.78-2.01]</td>
</tr>
<tr>
<td>FHS 2</td>
<td>1.52</td>
<td>[0.95-2.43]</td>
</tr>
<tr>
<td>FHS 3</td>
<td>1.81</td>
<td>[1.22-2.70]</td>
</tr>
<tr>
<td>FHS 4</td>
<td>1.18</td>
<td>[0.80-1.74]</td>
</tr>
<tr>
<td>FHS 5</td>
<td>1.14</td>
<td>[0.79-1.65]</td>
</tr>
<tr>
<td>FHS 6</td>
<td>1.12</td>
<td>[0.79-1.59]</td>
</tr>
<tr>
<td>Iceland</td>
<td>1.29</td>
<td>[1.06-1.57]</td>
</tr>
<tr>
<td>KORA S3</td>
<td>0.90</td>
<td>[0.70-1.16]</td>
</tr>
<tr>
<td>Maywood</td>
<td>0.88</td>
<td>[0.49-1.59]</td>
</tr>
<tr>
<td>Scandinavia</td>
<td>1.25</td>
<td>[0.69-2.24]</td>
</tr>
</tbody>
</table>
rs7566605 Genotype and BMI $\geq$ 30 kg/m$^2$ in Family Cohorts (Lyon et al, *PLoS Gen* 2007)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>C/C</th>
<th>C/G</th>
<th>G/G</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAMP</td>
<td>18.05</td>
<td>17.97</td>
<td>17.52</td>
<td>0.026</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>18.19</td>
<td>17.46</td>
<td>17.72</td>
<td>0.027</td>
</tr>
<tr>
<td>Scandinavia</td>
<td>25.70</td>
<td>26.43</td>
<td>26.43</td>
<td>0.96</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td>0.004</td>
</tr>
</tbody>
</table>
Possible Explanations of Heterogeneity of Results in Genetic Association Studies

- Biologic mechanisms
  - Genetic heterogeneity
  - Gene-gene interactions
  - Gene-environment interactions
- Spurious mechanisms
  - Selection bias
  - Information bias
  - Publication bias
  - Confounding (population stratification)
  - Cohort, age, period (secular) effects
  - Type I error
Possible Explanations of Heterogeneity of Results in Genetic Association Studies

• Biologic mechanisms
  – Genetic heterogeneity
  – Gene-gene interactions
  – Gene-environment interactions

• Spurious mechanisms
  – Selection bias
  – Information bias
  – Publication bias
  – Confounding (population stratification)
  – Cohort, age, period (secular) effects
  – Type I error
Possible Explanations of Heterogeneity of Results in Genetic Association Studies

• Biologic mechanisms
  – Genetic heterogeneity
  – Gene-gene interactions
  – Gene-environment interactions

• Spurious mechanisms
  – Selection bias
  – Information bias
  – Publication bias
  – Confounding (population stratification)
  – Cohort, age, period (secular) effects
  – Type I error
Possible Explanations of Heterogeneity of Results in Genetic Association Studies

• Biologic mechanisms
  – Genetic heterogeneity
  – Gene-gene interactions
  – Gene-environment interactions

• Spurious mechanisms
  – Selection bias
  – Information bias
  – Publication bias
  – Confounding (population stratification)
  – Cohort, age, period (secular) effects
  – Type I error
Definition of Bias

“Any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth.”

To be distinguished from random error…

Correlation between Discordance and Call Rate, Comparing ~250K SNPs in Common Autosomal SNPs, Call Rate > 90%

Courtesy, K. Doheny
Correlation between Discordance and Call Rate, Comparing ~250K SNPs in Common

Autosomal SNPs, Call Rate > 90%

Call Rate Genotyping Platform 1

Discordance Rate

Sample

0.0% 0.5% 1.0% 1.5% 2.0% 2.5%

96.5% 97.0% 97.5% 98.0% 98.5% 99.0% 99.5% 100.0%

1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86

Courtesy, K. Doheny
Correlation between Discordance and Call Rate, Comparing ~250K SNPs in Common

Autosomal SNPs, Call Rate > 90%

Sample

Call Rate

Discordance Rate

Call Rate Genotyping Platform 1
Call Rate Genotyping Platform 2
Discordance Rate

Courtesy, K. Doheny
Key Requirements for a Bias-Free Case-Control Study

• Cases are representative of all those in the study base who develop the disease
• Controls are representative of all those in the study base at risk of developing the disease and eligible to become cases and be detected in the study
• Collection of risk factor and exposure information is the same for cases and controls
• *Ancestral geographical origins and predominant environmental exposures of cases do not differ dramatically from controls*
Key Requirements for a Bias-Free Case-Control Study

- Cases are representative of all those in the study base who develop the disease.
- Controls are representative of all those in the study base at risk of developing the disease and eligible to become cases and be detected in the study.
- Collection of risk factor and exposure information is the same for cases and controls.
- Ancestral geographical origins and predominant environmental exposures of cases do not differ dramatically from controls.
Selection Bias: systematic differences between those who are selected for study and those who are not

• Prevalence-incidence or survival bias: Selection of currently available, existing cases will miss fatal and short episodes, and may miss mild or silent cases

• Non-response bias: Differential rates of non-response to inquiries between cases and controls

• Membership bias: Membership in a group (blood donors, Army recruits) may imply a degree of health differing systematically from the general population

• Referral or admission rate bias: Cases who are more likely to receive advanced treatment (those with greater access to health care or co-existing illness) may distort associations with other factors

Are cases representative of all those who develop the disease?

- To assess representativeness and potential biases, need to know how cases defined
- Study of atrial fibrillation (Gudbjartsson et al, 2007)
  - Sample 1: hospital diagnosis of AF “confirmed by 12-lead ECG”
  - Sample 2: Patients with ischemic stroke or TIA, diagnosis of AF “based on 12-lead ECG”
  - Sample 3: Patients hospitalized with acute stroke “diagnosed with AF”
  - Sample 4: Patients with lone AF or AF plus hypertension referred to arrhythmia service, “AF documented by ECG”

Are controls representative of disease-free persons eligible to become cases in the study?

• Also need to know how controls selected and determined to be disease-free
• Study of gallstones (Buch et al, 2007)
  – Sample 1: Gallstone-free controls from single hospital (vs 9 hospitals providing cases defined as post-cholecystectomy for cholelithiasis) from records of routine ultrasound US tests
  – Sample 2: Local population register undergoing additional exam with negative US
  – Sample 3: Population sample undergoing abdominal US to determine either “gallstone carrier status or previous hx cholecystectomy”

Information Bias: systematic differences in data collection or reporting between cases and controls

- Recall bias: Questions about specific exposures may be asked more frequently of cases, or cases may search their memories more intensively.
- Family information bias: The flow of family information about exposures or illnesses may be stimulated by, or directed to, a new case in its midst.
- Exposure suspicion bias: Knowledge of a patient’s disease status may influence the intensity and outcome of search for exposure to a putative cause.
- Instrument bias: Defects in calibration or maintenance of measurement instruments may lead to systematic deviations from true values.

Is risk factor information collected the same way in cases and controls?

- Cases of schizophrenia ascertained through local treatment facilities, physician referrals, advocacy groups, Web sites, media announcements and ads
  - Personal interview for psychotic, mood, and substance-use disorders, medical history
  - Family informant interview for patient history and family psychiatric history
- Controls recruited by random-digit dialing, completed preliminary consent and clinical assessment online
  - Screen for lifetime common mood, anxiety and substance use disorders
  - Lifetime psychosis, bipolar disorder, nicotine dependence, neuroticism and extraversion

Is DNA collected and handled the same way in cases and controls?

- 816 cases T1D from GRID study
- 877 controls from 1958 British Birth Cohort Study
- 6,322 nonsynonymous SNPs
- Samples from lymphoblastoid cell lines extracted using same protocol in two different labs
- Case and control DNAs arranged randomly, teams masked to case-control status
- Some extreme associations could not be replicated by second genotyping method
- Four rather than three data clouds for some nsSNPs

Signal Intensity Plots for \textit{CD44} SNP rs96666607

Information Bias: systematic difference in ancestral geographical origins and predominant environmental exposures between cases and controls

- Population structure: confounding by ancestral origin (stay tuned)
- Confounding by demographics or environmental exposures
Confounding

- Confounder: “A factor that distorts the apparent magnitude of the effect of a study factor on risk. Such a factor is a determinant of the outcome of interest and is unequally distributed among the exposed and the unexposed” (Last, 1983).
  - Associated with exposure
  - Independent cause or predictor of disease
  - Not an intermediate step in causal pathway

### FTO Variants, Type 2 Diabetes, and Obesity (Frayling 2007 and Zeggini 2007)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTCCC phase 1</td>
<td>1.27</td>
<td>[1.16-1.37]</td>
<td>$2 \times 10^{-8}$</td>
</tr>
<tr>
<td>WTCCC phase 2</td>
<td>1.22</td>
<td>[1.12-1.32]</td>
<td>$5 \times 10^{-7}$</td>
</tr>
<tr>
<td>DGI</td>
<td>1.03</td>
<td>[0.91-1.71]</td>
<td>0.25</td>
</tr>
</tbody>
</table>
# FTO Variants, Type 2 Diabetes, and Obesity (Frayling 2007 and Zeggini 2007)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Diabetes Association</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>WTCCC phase 1</td>
<td>1.27</td>
<td>[1.16-1.37]</td>
<td>$2 \times 10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>WTCCC phase 2</td>
<td>1.22</td>
<td>[1.12-1.32]</td>
<td>$5 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>DGI</td>
<td>1.03</td>
<td>[0.91-1.71]</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BMI Association (kg/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
</tr>
<tr>
<td>WTCCC Cases</td>
</tr>
<tr>
<td>WTCCC Controls</td>
</tr>
</tbody>
</table>
# FTO Variants, Type 2 Diabetes, and Obesity (Frayling 2007 and Zeggini 2007)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Diabetes Association</th>
<th>BMI Association (kg/m²)</th>
<th>Diabetes Association Adjusted for BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td>WTCCCC phase 1</td>
<td>1.27</td>
<td>[1.16-1.37]</td>
<td>2 x 10⁻⁸</td>
</tr>
<tr>
<td>WTCCCC phase 2</td>
<td>1.22</td>
<td>[1.12-1.32]</td>
<td>5 x 10⁻⁷</td>
</tr>
<tr>
<td>DGI</td>
<td>1.03</td>
<td>[0.91-1.71]</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Identifying Confounders

• Conduct literature review to ascertain currently known risk factors
• Collect data on known risk factors and other potential confounders
• Identify differences between cases and controls in prevalence of potential confounders: “Table 1,” comparing cases and controls, is crucial!
• Identify associations of potential confounders with risk factor of interest
• Adjust associations for confounders and compare estimates, look for ~10-20% difference

Distribution of Four Covariates in Case-Control Study of Nicotine Dependence

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Cases (n = 1,050)</th>
<th>Controls (n=879)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex (%)</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>38</td>
<td>37</td>
</tr>
<tr>
<td>Fagerström (score)</td>
<td>6.3</td>
<td>0</td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US (n)</td>
<td>797</td>
<td>713</td>
</tr>
<tr>
<td>Australia (n)</td>
<td>253</td>
<td>66</td>
</tr>
</tbody>
</table>

Do determinants of dependence differ in men and women?  
Do determinants of dependence differ in US and Australia?  

### Distribution of Three Covariates in Case-Control Study of Neovascular AMD

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Cases (n = 96)</th>
<th>Controls (n = 130)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male sex (%)</td>
<td>68</td>
<td>33</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>75</td>
<td>74</td>
</tr>
<tr>
<td>Smokers (%)</td>
<td>63</td>
<td>26</td>
</tr>
</tbody>
</table>

Do determinants of AMD differ in men and women?  
Do determinants of AMD differ in smokers and non-smokers?

Dealing with Confounders

• In design:
  – Randomize
  – Restrict: confine study subjects to those within specified category of confounder
  – Match: select cases and controls so confounders equally distributed

• In analysis:
  – Standardize: for age, gender, time
  – Stratify: separate sample into subsamples according to specified criteria (binning?)
  – Multivariate analysis: adjust for many confounders

Interaction: Is *LIPC* Genotype Related to HDL-C?

Inverse Relation between Endotoxin Exposure and Allergic Sensitization by CD14 Genotype

Dealing with Interaction

• Definition: differences in the association of one factor with a second factor according to the level of a third factor

• Beware: most studies are underpowered to identify interactions, formal interaction terms often not tested (Patsopoulous et al, *JAMA* 2007; 298:880-893)

• If it’s really there, rejoice!

• Stratify, do NOT adjust!

• May provide clues to biologic mechanisms
Replicating genotype–phenotype associations

What constitutes replication of a genotype–phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies
The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)\textsuperscript{1-3}. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even ‘agnostic’, approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype–phenotype associations, replication of which has often failed in independent studies\textsuperscript{4-7}. As the transition to genome-wide association studies occurs, the challenge will be to separate true associations from the blizzard of false positives attained through attempts to reproduce studies because of issues in either the initial study or the attempted replication\textsuperscript{4-6,20,37}. Small sample size is a frequent problem and can result in a conclusion from the literature because follow-up studies have not consistently analysed the same markers or those in perfect linkage dis-

Replication,
Replication, Replication,
Replication, Replication, Replication

Initial study:
- Sufficient description to permit replication
- Suggested criteria for soundness of initial report

Replication study:
- Similar population, similar phenotype
- Same genetic model, same SNP, same direction
- Adequately powered to detect postulated effect

Information to be Included in Initial Report

- Study information:
  - Source of cases and controls
  - Methods used for defining affection status
  - Participation rates and flow chart of selection
  - Standard “Table 1,” including rates of missing data
  - Success rate of DNA acquisition, comparability

- Genotyping and quality control procedures

- Results
  - Analysis methods in sufficient detail to understand and reproduce what was done
  - Simple single-locus and multi-marker (haplotype) association analyses
  - Significance of any known 'positive controls'

Why Medical Research Findings Are False

Summary

There is increasing evidence that most current published medical research findings are false. The probability that a result is true may be as low as 1 in 20 (false positives). Therefore, most published research findings should be treated with skepticism.

The number of true relationships that may be found increases exponentially with the number of tested relationships, whereas the number of false relationships decreases. The probability of finding a true relationship in a field is likely to be much greater than the probability of finding a false one. The number of true relationships can vary greatly depending on the number of tests conducted and the significance level used.

Conclusion

The number of true relationships that may be found increases exponentially with the number of tested relationships, whereas the number of false relationships decreases. The probability of finding a true relationship in a field is likely to be much greater than the probability of finding a false one. The number of true relationships can vary greatly depending on the number of tests conducted and the significance level used.
Controlling Bias in Genomic Research: Design

- Define population to be studied
- Maximize representativeness
- Use standard, reproducible methods for assignment of case/control status
- Use incident cases
- Select controls who are also eligible to be cases
- Estimate (and maximize!) participation rates
- Apply standard genotyping QC methods
- Replicate positive findings on different genotyping platform
Controlling Bias in Genomic Research: Analysis and Interpretation

- Describe sources of cases and controls
- Describe methods of disease ascertainment
- Compare participants and non-participants
- Compare cases and controls
- Stratify and adjust for important confounders (including population stratification)
- Stratify and test for important interactions
- Report results of genotyping QC
- Report results of prior known associations
“Yes ... I believe there’s a question there in the back.”

Reasonable Person Test: Does the Finding Make Sense?

• Bova et al studied MTHFR C677T variant in 48 persons with > 75% carotid stenosis compared to 26 persons with < 25% stenosis
• Persons with severe stenosis more likely to carry T allele
• Difference significant only in those with neither coronary nor peripheral arterial disease
• Carotid stenosis and coronary disease share major risk factors and are highly correlated

Is amyl nitrite associated with Kaposi’s sarcoma in homosexual men?

<table>
<thead>
<tr>
<th>Amyl Nitrite Use</th>
<th>Kaposi’s Sarcoma</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present</td>
<td>Absent</td>
<td></td>
</tr>
<tr>
<td></td>
<td>% (n/N)</td>
<td>% (n/N)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>60 (12/20)</td>
<td>15 (6/40)</td>
<td>8.5</td>
</tr>
<tr>
<td>Low</td>
<td>40 (8/20)</td>
<td>85 (34/40)</td>
<td>[2.4-29.6]</td>
</tr>
<tr>
<td>Total</td>
<td>100 (40/40)</td>
<td>100 (40/40)</td>
<td></td>
</tr>
</tbody>
</table>

Could an oncovirus explain some or all of the observed association?

<table>
<thead>
<tr>
<th>HIV Status</th>
<th>Amyl Nitrite Use</th>
<th>Kaposi’s Sarcoma</th>
<th>Odds Ratio [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>High</td>
<td>63 (12/19)</td>
<td>33 (3/9)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>37 (7/19)</td>
<td>67 (6/9)</td>
</tr>
<tr>
<td>Absent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>High</td>
<td>0 (0/1)</td>
<td>10 (3/31)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>100 (1/1)</td>
<td>90 (28/31)</td>
</tr>
</tbody>
</table>

Number of New, Significant Gene-Disease Associations by Year, 1984 - 2000

Of 600 Gene-Disease Associations, Only 6 Significant in > 75% of Identified Studies

<table>
<thead>
<tr>
<th>Disease/Trait</th>
<th>Gene</th>
<th>Polymorphism</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>DVT</td>
<td>F5</td>
<td>Arg506Gln</td>
<td>0.015</td>
</tr>
<tr>
<td>Graves’ Disease</td>
<td>CTLA4</td>
<td>Thr17Ala</td>
<td>0.62</td>
</tr>
<tr>
<td>Type 1 DM</td>
<td>INS</td>
<td>5’ VNTR</td>
<td>0.67</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>CCR5</td>
<td>32 bp Ins/Del</td>
<td>0.05-0.07</td>
</tr>
<tr>
<td>Alzheimer’s</td>
<td>APOE</td>
<td>Epsilon 2/3/4</td>
<td>0.16-0.24</td>
</tr>
<tr>
<td>Creutzfeldt-Jakob Disease</td>
<td>PRNP</td>
<td>Met129Val</td>
<td>0.37</td>
</tr>
</tbody>
</table>