National Human Genome Research Institute

National Institutes of Health

U.S. Department of Health and Human Services

Lecture 6: Bias in Human Genome Research

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
Senior Advisor to the Director, NHGRI,
for Population Genomics

July 18, 2008

Essay

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

PLoS Med. 2005 Aug;2(8):e124.

WSJ.com THE WALL STREET JOURNAL. ONLINE

SCIENCE JOURNAL

By ROBERT LEE HOTZ

Most Science Studies Appear to Be Tainted By Sloppy Analysis

September 14, 2007; Page B1

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.

WSJ. 2004Sep14.

Learning Objectives

- Appreciate impact of bias in genomic studies and how it differs from random error
- Define major types of bias in genetic association studies
- Identify sources of bias in published reports and potential impact on findings
- Avoid, reduce, or compensate for bias in one's own research

Impact of Bias on Association Results

- False negatives
- False positives
- Inaccurate effect sizes
  - Underestimates
  - Overestimates

Are there any other ways to be wrong??!

A Common Genetic Variant Is Associated with Adult and Ch

Alan H
Thoma
Grahar
Xiaofe
Nan M

Comment on "A Common Genetic Variant Is Associated with Adult an

Christ
Beatr

Comment on "A Common Genetic Variant Is Associated with Adult

Herbe
gene
to rep
gener
Dieter
Alexand
Heyo K.

Contrary
of the
increase
analysis
overwei

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

Ruth J. F. Loos,[1]* Inês Barroso,[2] Stephen O'Rahilly,[3] Nicholas J. Wareham[1]

Herbert *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.

- A
  /A
  re
- Asso
- Not s

Science, 14 Apr 2006

Science, 12 Jan 2007

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

Helen N. Lyon[1,2,3☯*], Valur Emilsson[4☯], Anke Hinney[5☯], Iris M. Heid[6,7☯], Jessica Lasky-Su[8,9☯], Xiaofeng Zhu[10☯], Gudmar Thorleifsson[4], Steinunn Gunnarsdottir[4], G. Bragi Walters[4], Unnur Thorsteinsdottir[4], Augustine Kong[4], Jeffrey Gulcher[4], Thuy Trang Nguyen[11,12], André Scherag[11,12], Arne Pfeufer[13,14], Thomas Meitinger[13,14], Günter Brönner[5], Winfried Rief[11,12], Manuel E. Soto-Quiros[15], Lydiana Avila[15], Barbara Klanderman[8], Benjamin A. Raby[8], Edwin K. Silverman[8], Scott T. Weiss[8], Nan Laird[8], Xiao Ding[8], Leif Groop[16,17,18], Tiinamaija Tuomi[17,18,19], Bo Isomaa[19], Kristina Bengtsson[17,18], Johannah L. Butler[1,2], Richard S. Cooper[10], Caroline S. Fox[21], Christopher J. O'Donnell[21], Caren Vollmert[6], Juan C. Celedón[8], H. Erich Wichmann[6,7], Johannes Hebebrand[5], Kari Stefansson[4], Christoph Lange[8], Joel N. Hirschhorn[1,2,22]

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.

Lyon HN et al, *PLoS Genet*; 2007 Apr 27;3(4):e61.

The Association of a SNP Upstream of *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

Lyon HN et al, *PLoS Genet*; 2007 Apr 27;3(4):e61.

rs7566605 C/C Genotype and BMI ≥ 30 kg/m^2 in Unrelated Individuals (Lyon et al, *PLoS Gen* 2007)

	Obesity Association			Frequency C/C	
Cohort	OR	95% CI	P-value	Cases	Controls
Essen	1.75	[1.15-2.67]	0.008	0.05	0.05
FHS 1	1.26	[0.78-2.01]	0.06	0.14	0.11
FHS 2	1.52	[0.95-2.43]	0.08	0.16	0.11
FHS 3	1.81	[1.22-2.70]	0.003	0.18	0.11
FHS 4	1.18	[0.80-1.74]	0.4	0.13	0.11
FHS 5	1.14	[0.79-1.65]	0.5	0.12	0.11
FHS 6	1.12	[0.79-1.59]	0.5	0.13	0.11
Iceland	1.29	[1.06-1.57]	0.007	0.13	0.11
KORA S3	0.90	[0.70-1.16]	0.4	0.10	0.11
Maywood	0.88	[0.49-1.59]	0.7	0.06	0.06
Scandinavia	1.25	[0.69-2.24]	0.5	0.13	0.10

rs7566605 Genotype and BMI ≥ 30 kg/m^2 in Family Cohorts (Lyon et al, *PLoS Gen* 2007)

	Mean Body Mass Index (kg/m^2)			
Cohort	C/C	C/G	G/G	P-value
CAMP	18.05	17.97	17.52	0.026
Costa Rica	18.19	17.46	17.72	0.027
Scandinavia	25.70	26.43	26.43	0.96
Combined				0.004

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

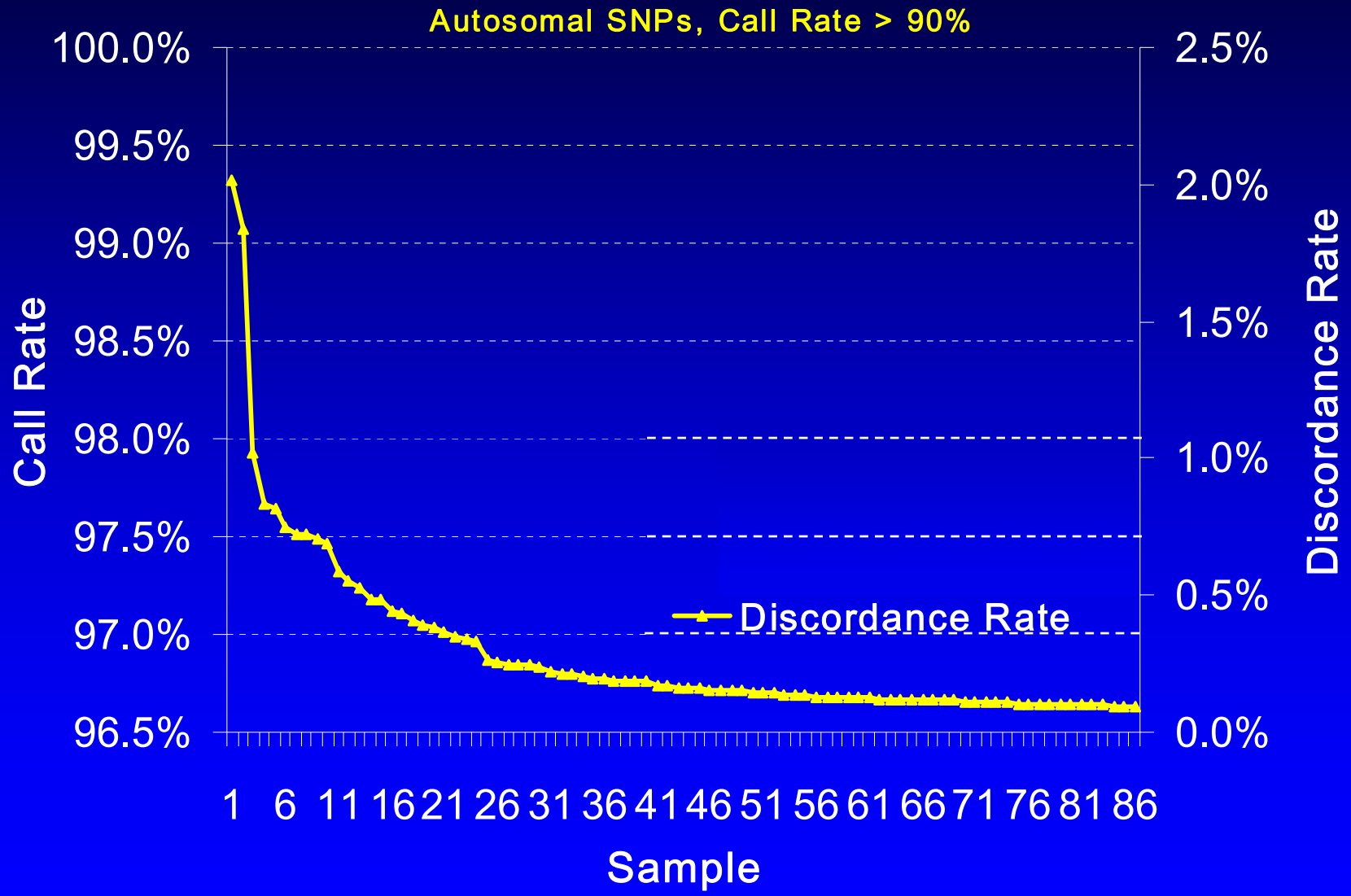
Definition of Bias

"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…

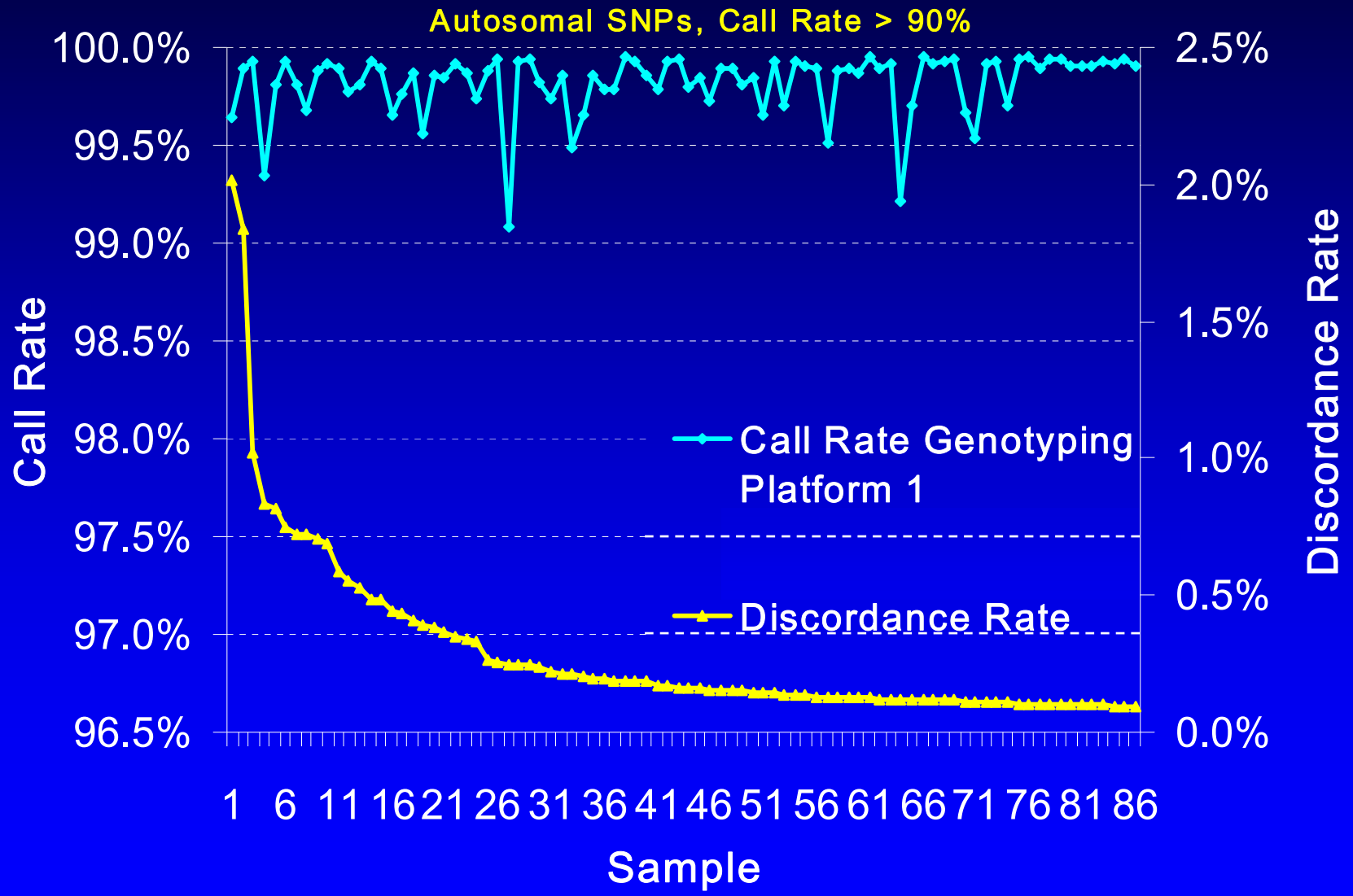
Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32:51-63.

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



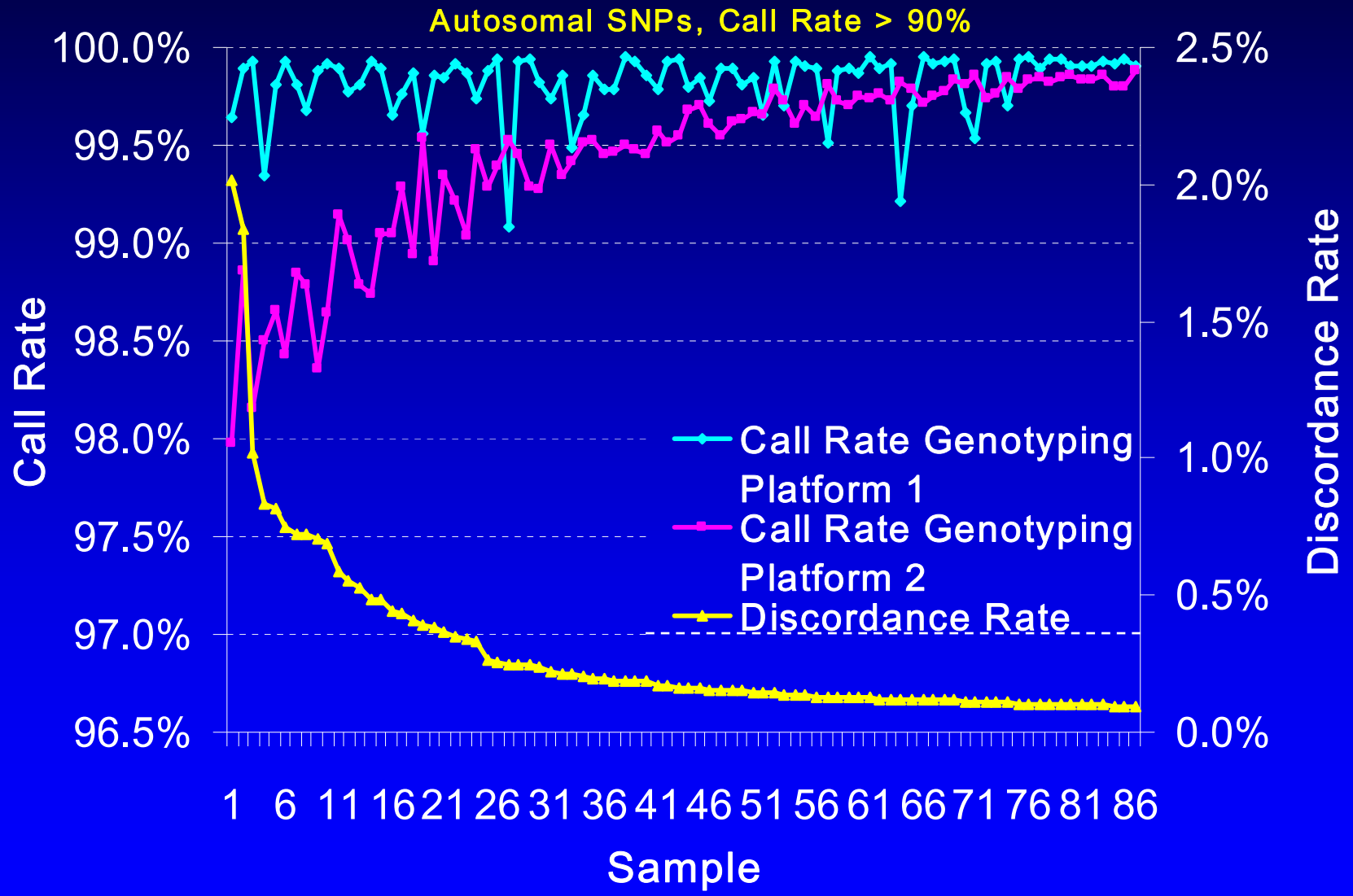
Courtesy, K. Doheny, CIDR

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



Courtesy, K. Doheny, CIDR

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



Courtesy, K. Doheny, CIDR

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

Selection Bias

- 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*

Information Bias

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

Sackett D, *J Chron Dis* 1979; 32:51-63 and Schlesselman J, *Case-Control Studies*, 1982.

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”

Gudbjartsson et al., *Nature* 2007; 448:353-57.

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”

Buch et al., *Nat Genet* 2007; 39:995-99.

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

Sackett D, *J Chron Dis* 1979; 32:51-63 and Schlesselman J, *Case-Control Studies*, 1982.

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

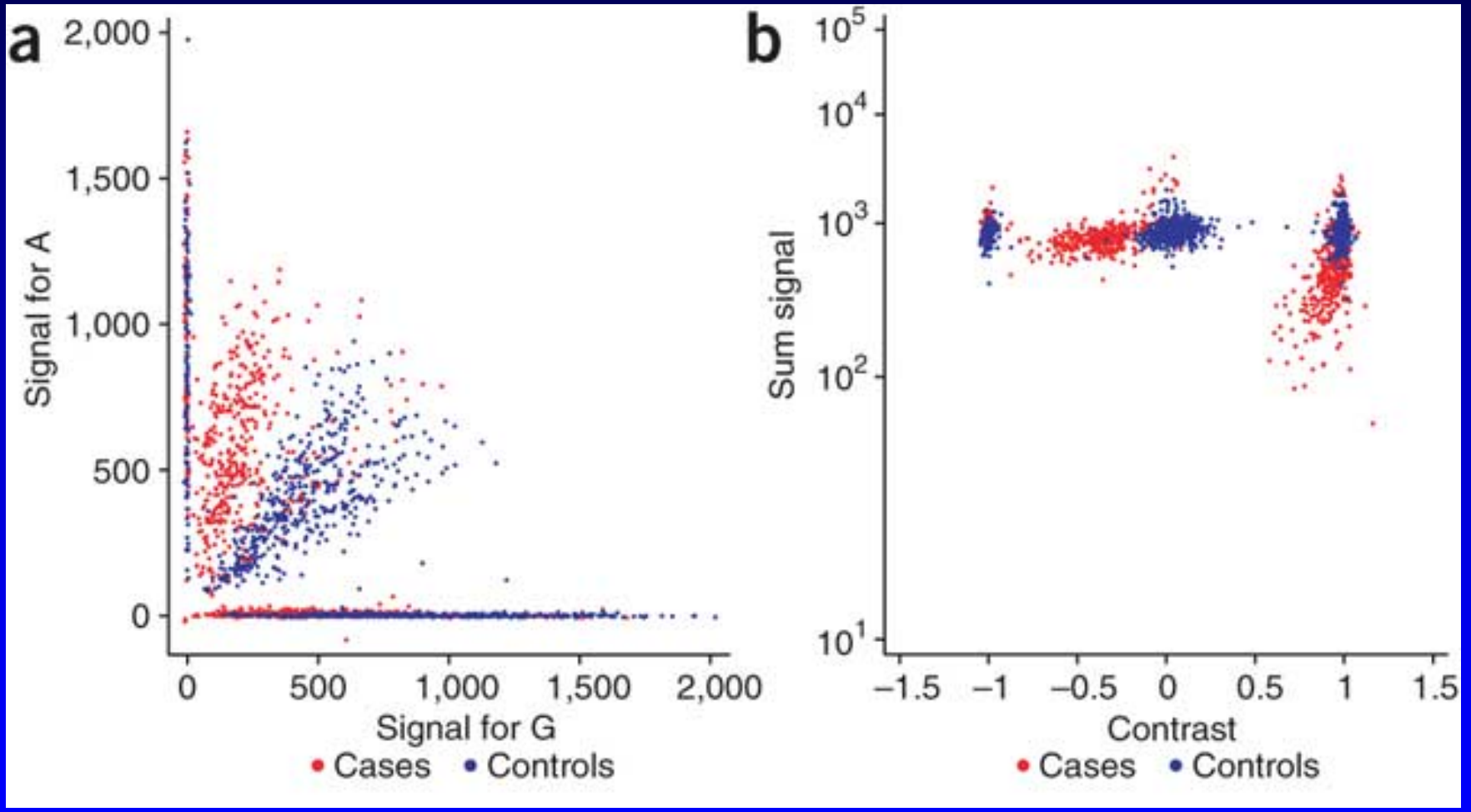
Suarez BK et al., *Am J Hum Genet* 2006; 78:315-33 and NIMH Genetics Initiative.

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

Clayton DG et al., *Nat Genet* 2005; 37:1243-46.

Signal Intensity Plots for *CD44* SNP rs9666607



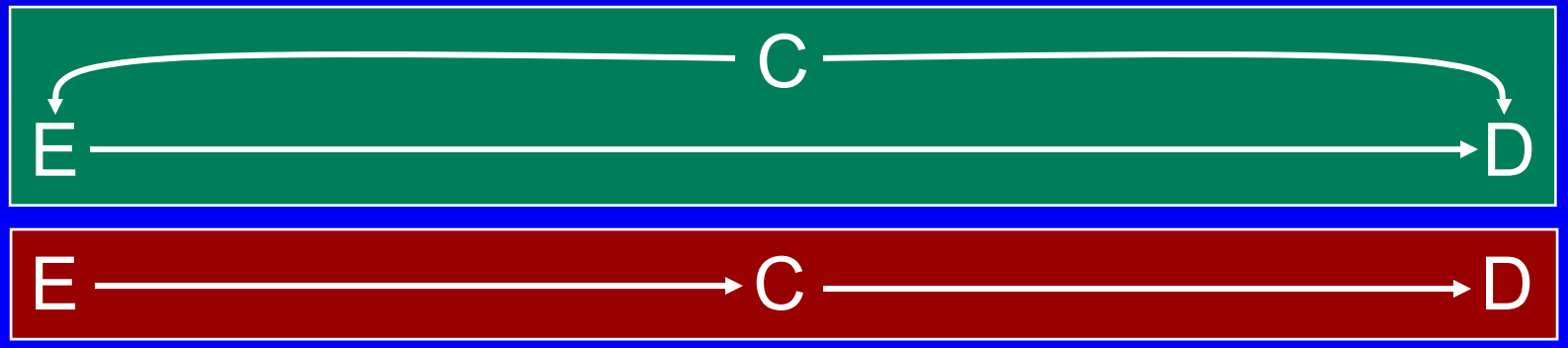
Clayton DG et al., *Nat Genet* 2005; 37:1243-46.

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

- Population structure: confounding by ancestral origin
- 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



Aschengrau and Seage, *Essentials of Epidemiology in Public Health,* 2003.

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

	Diabetes Association		
Cohort	OR	95% CI	P-value
WTCCC phase 1	1.27	[1.16-1.37]	2×10^{-8}
WTCCC phase 2	1.22	[1.12-1.32]	5×10^{-7}
DGI	1.03	[0.91-1.71]	0.25

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

	Diabetes Association		
Cohort	OR	95% CI	P-value
WTCCC phase 1	1.27	[1.16-1.37]	2×10^{-8}
WTCCC phase 2	1.22	[1.12-1.32]	5×10^{-7}
DGI	1.03	[0.91-1.71]	0.25

	BMI Association (kg/m^2)		
	TT	AT	AA
WTCCC Cases	30.2	30.5	32.0
WTCCC Controls	26.3	26.3	27.1

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

	Diabetes Association		
Cohort	OR	95% CI	P-value
WTCCC phase 1	1.27	[1.16-1.37]	2×10^{-8}
WTCCC phase 2	1.22	[1.12-1.32]	5×10^{-7}
DGI	1.03	[0.91-1.71]	0.25

	BMI Association (kg/m^2)		
	TT	AT	AA
WTCCC Cases	30.2	30.5	32.0
WTCCC Controls	26.3	26.3	27.1

	Diabetes Association Adjusted for BMI		
	OR	95% CI	P-value
WTCCC phase 2	1.03	[0.96-1.10]	0.44

But What About...

POPULATION STRATIFICATION!?!

Worldwide Genetic Variation in Frequency of Nine Alleles of D13S1493

World Average
(52 populations)

King MC, Motulsky AG, *Science* 2002; 298:2342-43.

Admixture and Population Stratification

- Admixture: matings among persons of two groups differing in their ancestral origins
- Each population has unique genetic and social history yielding differences in AF; ancestral patterns of migration, mating practices, reproductive expansions/bottlenecks, random variation
- Population stratification: differences in AF between cases and controls due to diversity in populations of origin and *unrelated to disease*
- Requires:
  1) population differences in disease prevalence
  2) population differences in allele frequencies

Cardon LR and Palmer LJ, *Lancet* 2003; 361:598-604.

Population Stratification and Allelic Association with Diabetes

Full heritage Am Indian population	European ancestry population
NIDDM prevalence: 40%	NIDDM prevalence: 15%

Population Stratification and Allelic Association with Diabetes

Full heritage Am Indian population	European ancestry population
NIDDM prevalence: 40%	NIDDM prevalence: 15%

Gm3;5,13,14 haplotype	NIDDM (+)	NIDDM (-)
+	8%	29%
-	92%	71%

OR 0.27 [0.18,0.40]

Population Stratification and Allelic Association with Diabetes

Full heritage Am Indian population
Gm3;5,13,14 prevalence: 1%
NIDDM prevalence: 40%

European ancestry population
Gm3;5,13,14 prevalence: 66%
NIDDM prevalence: 15%

Gm3;5,13,14 haplotype	NIDDM (+)	NIDDM (-)
+	8%	29%
-	92%	71%

OR 0.27
[0.18,0.40]

Population Stratification and Allelic Association with Diabetes

Full heritage Am Indian population	European ancestry population
Gm3;5,13,14 prevalence: 1%	**Gm3;5,13,14 prevalence: 66%**
NIDDM prevalence: 40%	NIDDM prevalence: 15%

Gm3;5,13,14 haplotype	NIDDM (+)	NIDDM (-)
+	8%	29%
-	92%	71%

OR 0.27 [0.18,0.40]

Index Indian heritage	Gm3;5,13,14 (+)	Gm3;5,13,14 (-)
0	18%	20%
4	28%	29%
8	36%	39%

Cardon, Palmer, *Lancet* 2003; 361:598-604, after Knowler et al 1988.

Use of Unlinked Genetic Markers to Detect Population Stratification

- When population stratification present, associations can be demonstrated between disease and arbitrary markers with no physical linkage
- Population stratification allows marker-allele frequencies to vary among population subgroups
- Disease more prevalent in one subpopulation will be associated with *any* alleles in high frequency in that subpopulation
- If population stratification exists, can often be detected by analysis of unlinked marker loci

Pritchard JK, Rosenberg NA, *Am J Hum Genet* 1999; 65:220-28.

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

Aschengrau and Seage, *Essentials of Epidemiology in Public Health,* 2003.

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

Covariate	Cases (n = 1,050)	Controls (n=879)
Male sex (%)	44	30
Age (yrs)	38	37
Fagerström (score)	6.3	0
Site		
US (n)	797	713
Australia (n)	253	66

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

Bierut LJ et al., *Hum Molec Genet* 2007; 16:24-35.

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

Covariate	Cases (n = 96)	Controls (n = 130)
Male sex (%)	68	33
Age (yrs)	75	74
Smokers (%)	63	26

Do determinants of AMD differ in men and women?

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

DeWan A et al., *Science* 2006; 314:989-92.

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

Aschengrau and Seage, *Essentials of Epidemiology in Public Health,* 2003.

Essay

Why Mo…hed Research Fi…gs Are Fa…

John P. A. Ioanni…

Summary

There is incr…m that most current publish…findings are false. The prob…esearch claim is true may dep… power and bias, the numbe…dies on the same question, a…tly, the ratio of true to no rela…ng the relationships prob…entific field. In this framew…finding is less likely to be tru…dies conducted in a field ar… effect sizes are smaller; w… greater number and lesser … of tested relationships; where… greater flexibility in designs, def… outcomes, and analytical modes; wh…

fa…ce this problem and some…of.

Modeling…k for False Positive Fin…

Several methodo… pointed out [9–11]… rate of nonreplicatio… confirmation) of researc… is a consequence of the co… yet ill-founded strategy of clai… conclusive research findings sole… the basis of a single study assessed b… formal statistical significance, typically … for a *p*-value less than 0.05. Research … not most appropriately represente… …arized by *p*-values…

is character…eld and can vary a lot de…whether the field targets …relationships or searches f…or a few true relation…thousands and millions…es that may be postulat…o consider, for compu…licity, circumscr…here either there is only o…onship (among …othesized) or …to find any of the …e relationships. The …bility of a relationship …/(*R* + 1). The probability …nding a true relationship …e power 1 – β (one minus …pe II error rate). The probability

Ioannidis J, *PLoS Med.* 2005 Aug;2(8):e124.

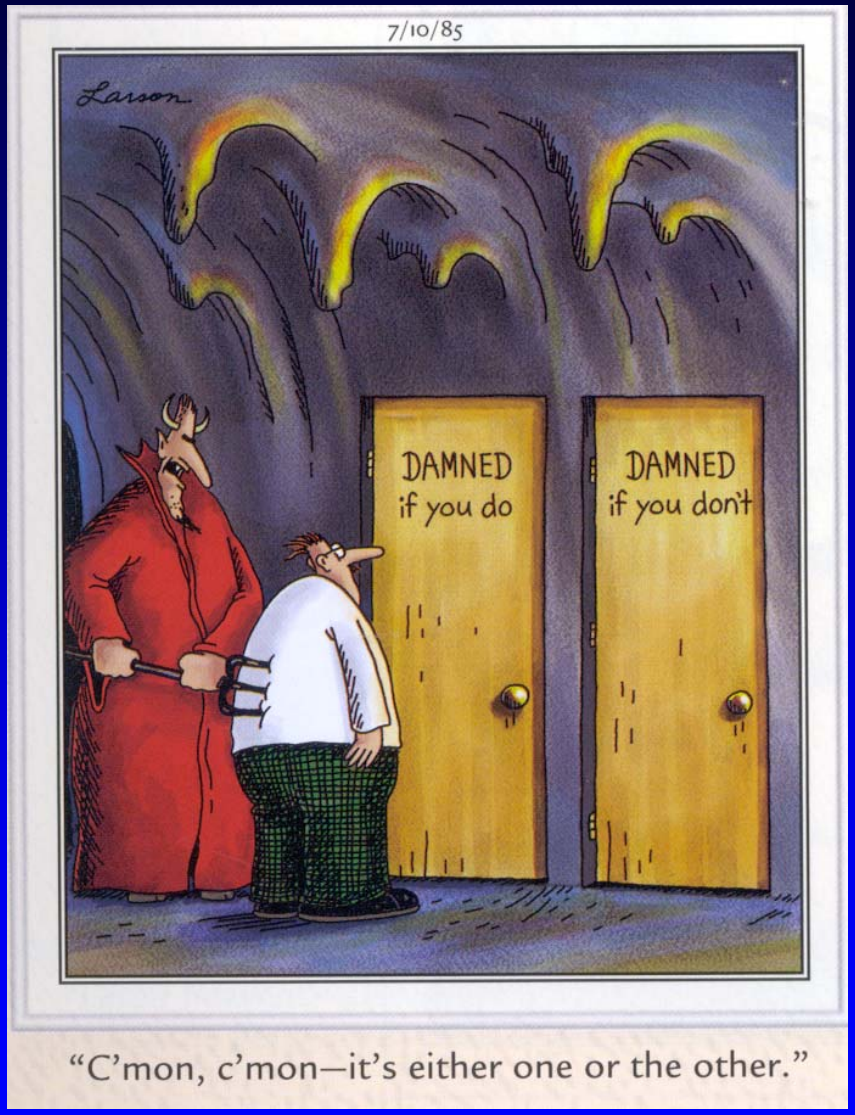
Controlling Bias in Genomic Research: Design

- Define population to be studied
- Maximize representativeness
- Use standard, reproducible meth
  assignment of case/control st
- Use incident cases
- Select control
  be
- Es
- 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

Class Participation Exercise!



Larson, G. *The Complete Far Side*. 2003.

- Dr. Y wants to know if the -514(C/T) *LIPC* polymorphism is related to HDL-cholesterol levels, so he genotypes the variant in 1,000 participants of the Framingham Study
- He finds that mean HDL-C levels are the same in persons with CC, CT, and TT genotypes and dejectedly publishes his findings in *J Negat Res*
- Three months later the lead story in *Nature* demonstrates that the C allele raises HDL-C in a dominant fashion in 200 community-dwelling Finns and calls into question Dr. Y's genotyping and/or phenotyping abilities…
- A month after that Dr. Z demonstrates in 2,000 Costa Ricans that it's actually the T allele that raises HDL-C, and in a dose-dependent (co-dominant) fashion, but no journal will accept her report. She writes a letter to the editor accusing her competitors of suppressing her research and of blocking her from getting peer-reviewed funding because she's a philatelist

What kind of biases could be operative here?

A. Selection bias

B. Publication bias

C. Incidence/prevalence bias

D. Instrument (measurement) bias

E. Membership bias

Two months later, while driving his taxi, Dr. Y sees a report from his former Framingham colleagues demonstrating that increased dietary fat intake raises HDL-C levels. He begins to wonder if differences in fat intake could be confounding the relationship of the -514(C/T) polymorphism and HDL-C levels.

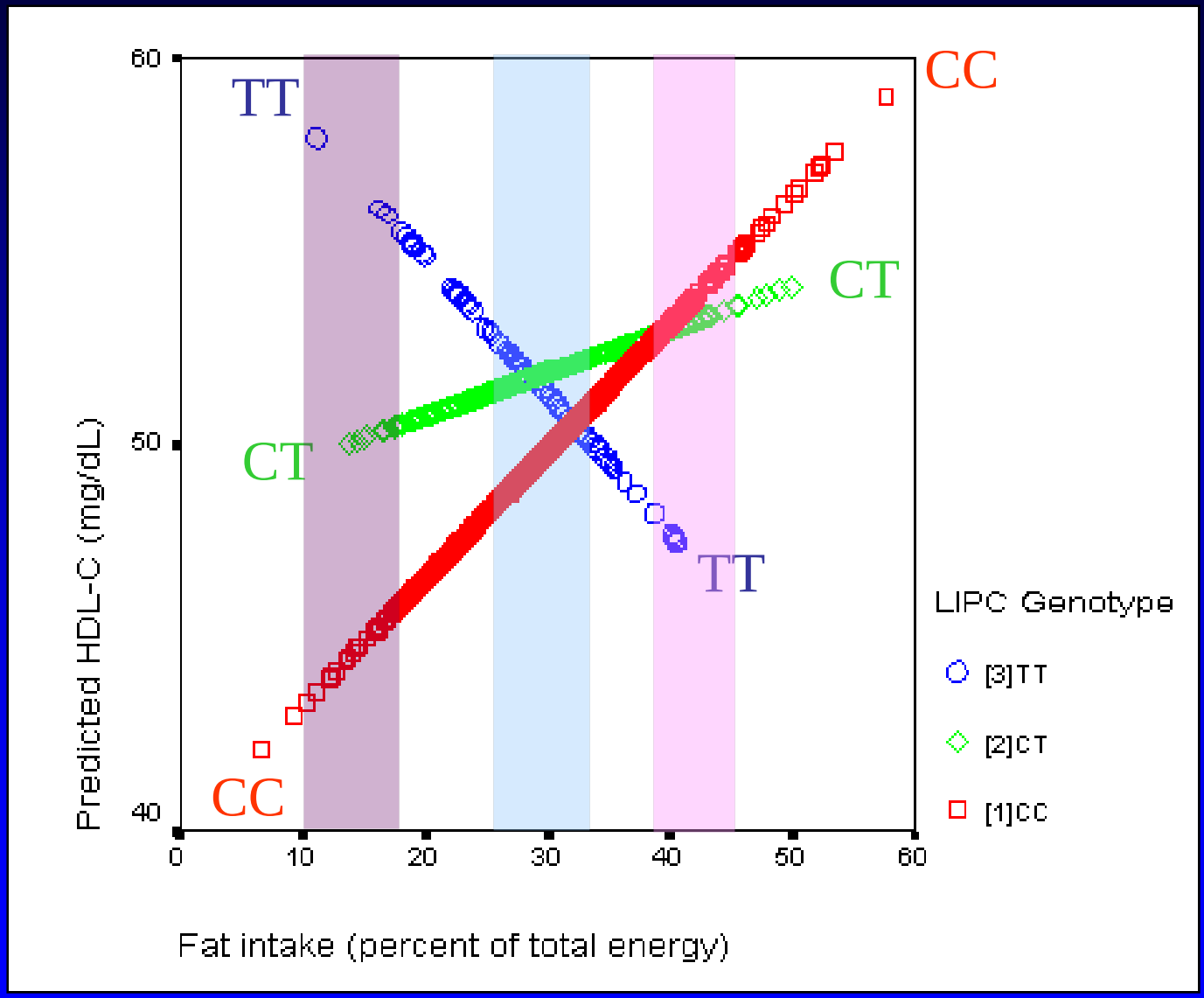
Which of the following would be evidence for possible confounding?

A. Fat intake differs significantly among Finns, Americans, and Costa Ricans

B. -514(C/T) frequencies differ significantly among the three groups

C. Feeding TT homozygotes nothing but cheeseburgers for 3 weeks makes their HDL levels fall

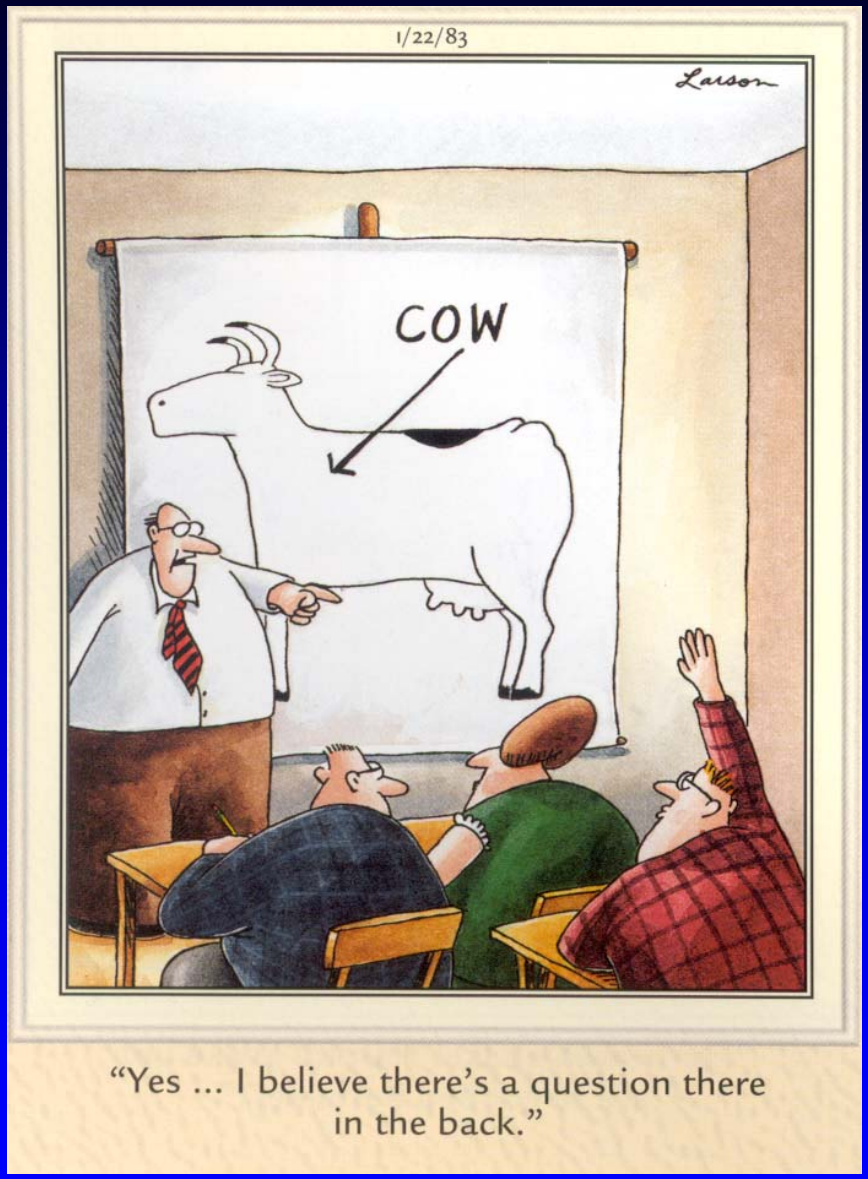
Interaction: Is *LIPC* Genotype Related to HDL-C?



Ordovas et al., *Circulation* 2002; 106:2315-21.

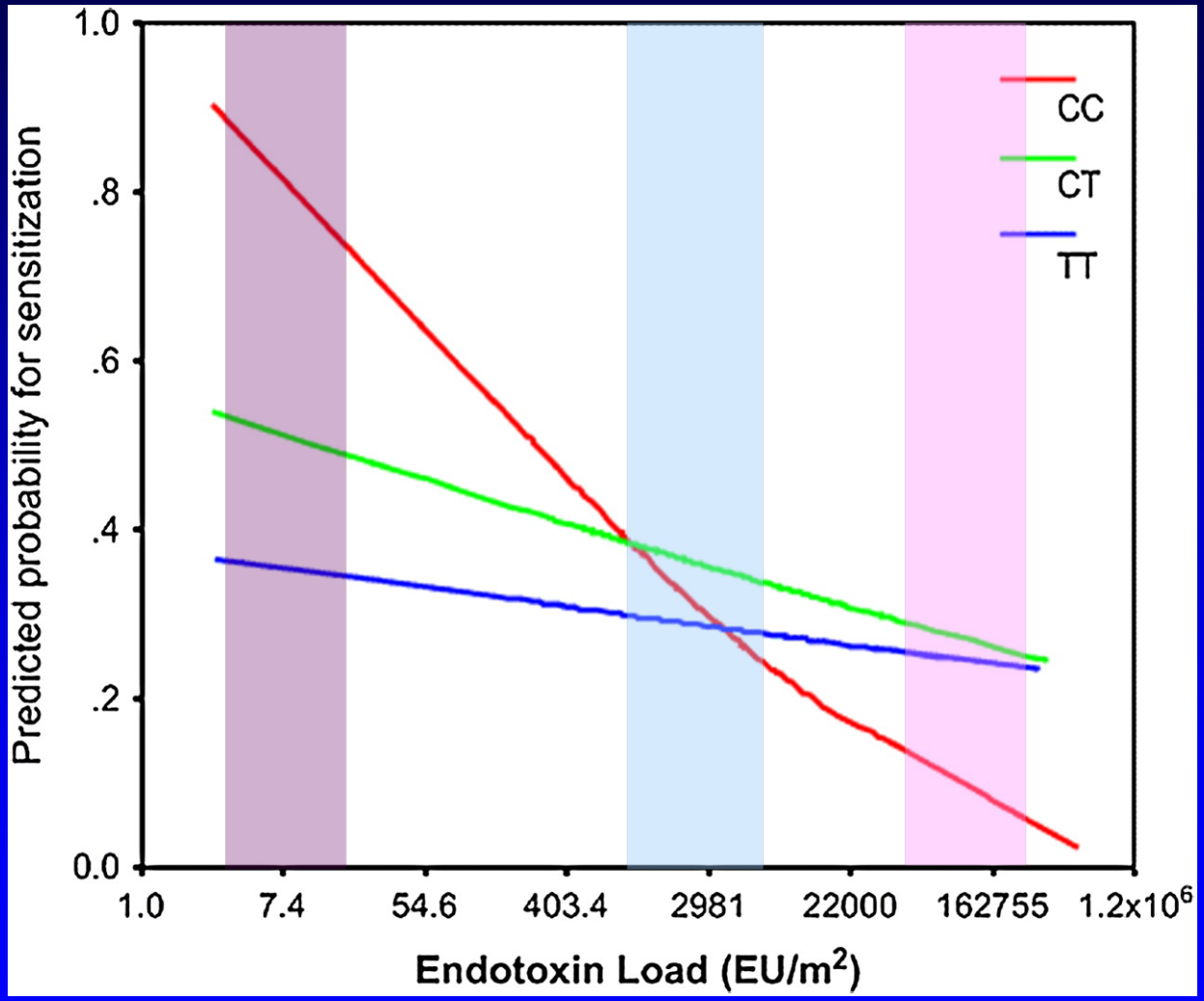
Dealing with Interaction

- 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 (Patsopoulos et al, *JAMA* 2007; 298:880-893)
- If it's really there, rejoice!
- Stratify, do NOT adjust!
- May provide clues to biologic mechanisms



Larson, G. *The Complete Far Side*. 2003.

Inverse Relation between Endotoxin Exposure and Allergic Sensitization by CD14 Genotype



Simpson A et al., *Am J Respir Crit Care Med* 2006;174:386-92.