AN ANALYSIS OF GENOME-WIDE ASSOCIATION STUDIES TO PRODUCE EVIDENCE USEFUL IN GUIDING THEIR REPORTING AND SYNTHESIS

ALEXANDER JOHN YURKIEWICH

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the MSc degree in Epidemiology

Epidemiology and Community Medicine
Faculty of Medicine
University of Ottawa

Student	Advisory Committee
Alexander John Yurkiewich	Dr. Julian Little
	Dr. David Moher

ABSTRACT

Introduction The present study evaluated reported methodological characteristics of GWAS, investigating relationships between reported methodological characteristics and outcomes observed.
Methods GWAS were identified from NHGRI's catalogue of GWAS (2005 to 2009). Multivariate meta-regression models (random effects) were produced to identify the impact of reported study characteristics and the strength of relationships between the variables and outcomes.
Results The summary odds ratios for replication components of GWAS in cancer was 1.34 (95% CI 1.25, 1.43) and neuropsychiatric disorders was 1.43 (95% CI 1.30, 1.57). Heterogeneity was accounted for by nature of the control group, relationship between case/control groups, whether cases/controls were drawn from the same population, if data was a primary collection or a build on pre-existing data, if quality assurance was reported, and if the study reported power/sample size.
Conclusion Evidence supports the existence of variability in reporting, with index components demonstrating less variability than replication components in the GWAS.

ACKNOWLEDGEMENTS

I would like to extend my gratitude to the following people, without whom the production of this thesis would not have been possible:

Dr. Julian Little, for the limitless advice and encouragement he gave throughout this entire process, from start to finish; and the patience and accommodation he was able to show me when I needed it most.

Dr. David Moher, for his insight and guidance during the production of this thesis.

Dr. Badr Aljasir, for the opportunity to work with him closely not only during the production of this thesis, but also with the production of his own.

Diana Lynn-Fox and Valery L'Heureux, for always providing assistance when it was needed.

Fay Draper and Paul-Andre David, for their attentiveness to the details of the program, and the constant communications they provided me.

My family, whom without I would not be who I am. It is their love and support that has allowed this process to take place.

TABLE OF CONTENTS

BACKGROUND 8

AIMS AND OBJECTIVES

Aim 15
Objectives 15

METHODS

Comparison of Reported Methodological Characteristics Over Time
Selection of Studies 16
Statistical Analysis 16

Genome-Wide Association Study Analysis
Selection of Studies 17
Outcome Measure 18
Language Restrictions 19
Power Calculation 19
Pilot Study 20
Data Extraction 27
Data Management 27
Statistical Analysis 28
Production of Guidelines 29

SIGNIFICANCE OF THESIS 30

RESULTS

Section 1: Comparison of Reported Methodological Characteristics Over Time 31

Section 2: Review and Analysis of Reporting of Genome-Wide Association Studies
2.1 Review and Analysis of Reporting of Index Components of Cancer 34
Description of Statistics 34
Meta-Analysis 40
2.2 Review and Analysis of Reporting of Index Components of Neuropsychiatric Disorders 58
Description of Statistics 58
Meta-Analysis 64
2.3 Review and Analysis of Reporting of Replication Components of Cancer 82
Description of Statistics 82
Meta-Analysis 85
2.4 Review and Analysis of Reporting of Replication Components of Neuropsychiatric Disorders 99
Descriptive Statistics 99
Meta-Analysis 102
2.5 Study Size Effects 117

DISCUSSION 122

CONCLUSION 132

REFERENCES 134

LIST OF TABLES

Table 1: Data Extraction Form Developed for Candidate Gene Association Studies[39] **21**

Table 2: Final Data Extraction Form **24**

Table 3: List of Additions and Deletions Made During Development of the Data Extraction Form **26**

Table 4: Comparison of the Methodological Characteristics of HuGE Journal Articles from Two Time Periods (2001-2003 vs. 2007) **32**

Table 5: Comparison of the Methodological Characteristics of HuGE Journal Articles from Two Time Periods (2006 vs. 2007) **33**

Table 6: Design and Conduct of Index Components of Genome-Wide Association Studies in Cancer **37**

Table 7: Analysis of Index Components of Genome-Wide Association Studies in Cancer **39**

Table 8: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Index Components of Genome-Wide Association Studies in Cancer **40**

Table 9: Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Discrete Variables): 40 studies **43**

Table 10: Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Continuous Variables) **53**

Table 11: Design and Conduct of Index Components of Genome-Wide Association Studies in Neuropsychiatric Disorders **61**

Table 12: Analysis of Index Components of Genome-Wide Association Studies in Neuropsychiatric Disorders **63**

Table 13: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders **65**

Table 14: Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables): 36 studies **68**

Table 15: Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables) **77**

Table 16: Design and Conduct of Replication Components of Genome-Wide Association Studies in Cancer **84**

Table 17: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Replication Components of Genome-wide Association Studies in Cancer **85**

Table 18: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Discrete Variables): 41 studies **88**

Table 19: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Continuous Variables) **96**

Table 20: Design and Conduct of Replication Components of Genome-Wide Association Studies in Neuropsychiatric Disorders **101**

Table 21: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders **103**

Table 22: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables): 20 studies **106**

Table 23: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables) **114**

Table 24: Results of Tests for Publication Bias in Index and Replication Components of Genome-wide Association Studies in Cancer and Neuropsychiatric Disorders **121**

LIST OF FIGURES

Figure 1: Types of Cancers Included in the Study Sample (n=53) **35**

Figure 2: Forest Plots of Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Discrete Variables) **45**

Figure 3: Scatter Plots of Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Continuous Variables) **54**

Figure 4: Neuropsychiatric Disorders Included in the Study Sample (n=74) **59**

Figure 5: Forest Plots of Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables) **70**

Figure 6: Scatter Plots of Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables) **78**

Figure 7: Forest Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Discrete Variables) **90**

Figure 8: Scatter Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Continuous Variables) **97**

Figure 9: Forest Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables) **108**

Figure 10: Scatter Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables) **115**

Figure 11: Assessment of Publication Bias by Funnel Plot of Precision by Log odds ratio of the Replication Components of Genome-wide Association Studies in Cancer **117**

Figure 12: Assessment of Publication Bias by Funnel Plot of Precision by Log odds ratio of the Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders **118**

Figure 13: Assessment of Publication Bias by Funnel Plot of Precision by Stouffer's Z-score of the Index Components of Genome-Wide Association Studies in Cancer **119**

Figure 14: Assessment of Publication Bias by Funnel Plot of Precision by Stouffer's Z-score of the Index Components of Genome-Wide Association Studies in Neuropsychiatric Disorders **120**

BACKGROUND

The first complete draft of the Human Genome Project was published in 2003, a major milestone in the history of human biology.[1] Human genome epidemiology, a fairly new branch of epidemiology concerned with the use of epidemiological methods to determine the impact of genetic variation in humans on the occurrence and progression of disease, has developed in the wake of this milestone and advances in related technologies.[2] Thus, scientists have been stimulated to develop new and innovative techniques for studying disease etiology with a long term view to prevent disease or improve treatment, and hence improve public health.[3] Using these techniques, scientists worldwide have been able to make new discoveries in the field of human genome epidemiology.

The field is fast-moving, both in terms of the volume of search and technological development (e.g., array technologies capable of detecting thousands of single nucleotide variations simultaneously, techniques for identifying copy number variations, and high throughput sequencing). However, the experience of investigation of candidate genes, in which the genes and variants have been selected for investigation based on biological and physiological information regarding the involvement of gene products in early developmental pathways, biochemical and cellular processes of progression, and/or clinical manifestations, has generated concern about the ability to make sense of these new discoveries. A lack of replication of studies of the association between candidate genes and diseases or traits has been recognized as a serious problem[20-22, 40-45] (not unique to genetic association studies[46,47]). Methodological problems may account for this,[1,10,48] but assessment is limited because of inadequate reporting.

The number of publications on the associations between genes and diseases has increased tremendously; with more than 60,000 published journal articles, the annual number has more than tripled between 2001 and 2011.[30,49] However, to date, there has been relatively limited assessment of

reporting practices. An assessment of a random sample of 315 genetic association studies published between 2001 and 2003 found that most studies provided some qualitative descriptions of the study participants (e.g., origin and enrolment criteria), but reporting of quantitative descriptors (e.g., age and sex) was variable.[38] Additionally, completeness of reporting of methods that allow readers to assess potential biases (e.g., number of exclusions or number of samples that could not be genotyped) varied.[38] Only some studies described methods to validate genotyping or mentioned whether research staff were blinded to outcome.[38] The same problems persisted in a smaller sample of studies published in 2006.[38]

Transparent reporting of human genome epidemiology studies is a necessity for genomic information to be useful in medicine and public health, since incomplete reporting of findings makes it difficult to assess the risk of bias. To promote transparent reporting, international collaborative efforts such as the production of the STREGA (STrengthening the REporting of Genetic Association studies) Statement,[5] have begun to make clear the processes by which all types of scientific research (including experimental studies, observational studies, systematic reviews and meta-analyses, etc.) are produced. A need to look at the baseline of reporting of candidate gene studies has been identified, primarily in order to assess the impact of the STREGA Statement.[5] Since there has been a lot of activity in general, obtaining a picture of this activity over time is now relevant. This is done with the hope that all scientific publications will one day include the information necessary to enable the risk of bias to be assessed, and to improve the quality of the overall body of literature.

Scientific journals now endorse various ways to ensure that transparent reporting has taken place, and the EQUATOR Network has been created in order to collect and make available all types of reporting guidelines necessary for transparent reporting of research.[6] Some important reporting guidelines currently being used include the CONSORT (CONsolidated Standards Of Reporting Trials) Statement for randomized controlled trials, the STROBE (STrenghthening the Reporting of Observational

studies in Epidemiology) Statement for observational studies, and the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) Statement for systematic reviews.[7-9]

Recently, epidemiologists have begun to come together and collaborate on projects designed to enhance the efficiency and reliability of the field of human genome epidemiology. One such group has proposed a Human Genome Epidemiology Road Map.[10] The Public Population Project in Genomics (P^3G), comprising members from 25 countries, was created in order to bring together various aspects of human genome epidemiology, including methods and tools, database information, and to promote compatibility between the components of the infrastructure of human genome epidemiology.[11]

A particularly important recent development is the genome-wide association study. The National Institutes of Health (NIH) have defined a genome-wide association study as a study of common genetic variation across the entire human genome designed to identify genetic associations with observable traits.[12] Unlike case-control studies investigating candidate gene-disease associations, which investigate the relationship between one or more specific gene variants postulated a priori to be associated with a disease, genome-wide association studies use very high-throughput methods to study many thousands of gene variants at once and relate them to the health-related characteristic in question.[13] This new technique clearly offers enormous potential for human genome epidemiology, making it possible to investigate potential relationships between gene variants related to health-related characteristics at a much higher rate than by the hypothesis-based approach of testing only candidate variants of candidate genes.[14] According to the National Human Genome Research Institute's (NHGRI) catalogue of published genome-wide association studies,[4] upwards of 4,500 single-nucleotide polymorphisms (SNPs) related to over 500 different diseases/traits have been identified and replicated in genome-wide association studies. Moreover, as of January 22, 2009, an open access database of

genome-wide association study results containing 56,411 significant SNP-phenotype associations and accompanying information from 118 genome-wide association studies has been created.[28]

Although the benefits of genome-wide association studies are clearly evident, they are not without their drawbacks. One issue with the approach of a genome-wide association study is that the multiplicity of statistical tests required to perform a genome-wide association study introduces the possibility for false-positive results, necessitating the redefinition of acceptable levels of significance and emphasizing the importance of replication of findings.[15] Due to the wide variety of genome-wide association studies now being produced, and their potential, it is becoming increasingly important to find a way to assess the risk of bias in the mosaic of evidence available by understanding its patterns.[16]

A second issue regarding genome-wide association studies relates to linkage disequilibrium. Linkage disequilibrium is defined as the departure of the recombination rate of genetic markers in a population from chance expectation. Linkage disequilibrium can be influenced by a number of factors, including population structure, genetic linkage, recombination and mutation rates, selection, or genetic drift.[233] In most populations, there are established patterns of linkage disequilibrium. Genome-wide association studies make use of linkage disequilibrium, in that the SNPs genotyped directly may also act as markers for association with other ungenotyped genetic variants in the same region of a gene.[17,50] Several projects, notably the International HapMap Project, have classified genetic sequences of reference individuals from diverse populations in order to identify regions of the genome where genetic variants are identical to one another.[27] In certain situations, because of the breakdown of established patterns of linkage disequilibrium, genome-wide association studies become vulnerable to a higher prevalence of false positive results. In order to combat this, rather than attempting to solve this issue by way of Bonferroni adjustment for multiple statistical comparisons, Bayesian inference or various exploratory methods, researchers utilizing genome-wide association studies are being directed to

perform studies with multiple replications, in addition to replicating previously published studies, so a systematic attempt can be made at integrating and interpreting the entire scope of epidemiological data that genome-wide association studies have begun to produce.[235-238]

The final issue with the utilization of genome-wide association studies stem from their hypothesis-free approach. Although genome-wide association studies are highly successful for identifying the relationship between genetic variation and disease, often the pathology of disease cannot be explained by single genetic effects. Gene-environment interactions[18] and the identification of metabolic pathways[19] responsible for more complex pathologies (e.g., inflammatory bowel disease (IBD)) are two important characteristics in the development of disease that cannot be identified with the use of single genome-wide association studies. These more complex pathologies arise from a chain of events which occur over the course of a metabolic pathway, and involve the interaction of many different genes and the environment. Despite this, the genome-wide association study can contribute as a powerful approach for identifying single gene effects within the context of more complex diseases.

In the past, the value of research that has been performed in various fields has become difficult to assess due to lack of transparency resulting from incomplete reporting of methods and poor reporting of results.[23-25] Although the STREGA Statement demonstrates a small amount of relevance towards genome-wide association studies, not enough information has been published about the potential for bias in order to fully address the STREGA Statement in the context of genome-wide association studies. Due to this, there is a definite need to evaluate the current state of knowledge by reviewing the reporting of a sample of genome-wide association studies.

Therefore, the purpose of this research is to perform a large-scale analysis of the reporting of genome-wide association studies in order to produce evidence useful in guiding their reporting and synthesis. The current investigator is confident that a review of this kind has not been performed.

Given that the outcomes of many of the genome-wide association studies included in this research will be different, it is important to consider that the "true" estimates of effect will differ across the group of studies. In order to account for this issue, one must understand that the presence of a "true" association between gene variant and disease varies between different gene variants and different diseases. Determining whether or not the differences in effect estimates are due to methodological characteristics or due to real differences with different outcomes becomes a concern. To date, however, only a very small proportion of gene-disease association studies investigated have been replicated.[20-22] The question that then remains is this: Do the "true" associations account for the differences in effect estimates across studies, or is it more likely that the differences in methodology contribute to the differences in outcome? To investigate this point statistically, different disease groups will be separated and analyzed individually in order to determine if the methodological rigor differs between groups of investigators working with different types of diseases.

This research will begin with an analysis of time trends identified from information collected in two previous independent analyses of case-control studies investigating gene-disease associations in order to examine if the reporting observed in these case-control studies has changed over time, to help determine whether any differences in reporting practices between candidate gene and genome-wide association studies might be due to general trends in reporting practices or are specific to study design. Following this, a large-scale analysis of genome-wide association studies will be performed in order to analyze their methodological characteristics.

Publication of research results has always been subject to reporting bias, a phenomenon which tends to alter the frequency in which all scientific results are presented to the scientific community at large. In his *Novum* Organon,[221]Francis Bacon wrote:

> "It is the peculiar and perpetual error of the human understanding to be more moved and excited by affirmatives than by negatives."

Various issues connected with selective reporting and publication bias have been documented, including under-reporting of insignificant or negative results,[204] over-reporting of significant positive results,[205-207] the nature and direction of the research,[208] speed of publication,[209-211] language of publication,[212,213] and influence on the citations of research in future journal articles.[214-219] Although it has been argued that genome-wide association studies might be less susceptible to selective reporting and publication bias, on the basis that they could be used to collect and interpret information on large collections of genetic variants simultaneously while concurrently making the information transparent and available in an online public database,[235,239] empirical evidence on the reporting of genome-wide association studies is lacking. Therefore, the purpose of this research is to examine the reporting of such studies. This research should help to guide those reporting future studies, and thereby assist in clarifying the risk for bias in genome-wide association studies.

In order to achieve this, the current investigator chose to utilize information from a few published sources for the purposes of creating a data extraction form suited for extracting relevant information from genome-wide association studies. These sources include the STREGA Statement,[5] journal articles published by McCarthy et al[231] and Manolio et al,[31] and the data extraction form created by Aljasir.[39] In particular, the STREGA Statement identified characteristics unique to genetic association studies that, when included, contribute to the overall transparency of reporting. These include, but are not limited to, population stratification, Hardy-Weinberg equilibrium, replication, selection of participants, statistical methods (e.g., power, sample size, heterogeneity), relatedness, and reporting of descriptive and outcome data. These sources formed a solid basis upon which a unique data extraction form could be created.

AIMS AND OBJECTIVES

Aim: To evaluate the reported methodological characteristics of genome-wide association studies, assess these in the broader context of reporting of genetic association studies, and investigate the possible relationship between reported methodological characteristics and the associations observed.

Objectives:

a) To compare information collected from three independent samples of case-control studies investigating gene-disease associations (years of publication: 2001-2003, 2006 and 2007) to identify if any changes in the results of the extracted information have occurred over time.

b) To identify a sample of genome-wide association studies and to extract various methodological characteristics and related information from the sample of genome-wide association studies, in order to assess if methodological characteristics are transparently reported, for the purposes of creating a possible adjunct to the STREGA (STrenghtening the REporting of Genetic Association studies) Statement regarding the reporting of genome-wide association studies.

c) To conduct meta-regression, by methodological characteristic, and multivariate analyses, to study the association between methodological characteristics and effect estimates, for both the primary genome-wide association study and, if applicable, the replication components embedded within the primary genome-wide association study.

METHODS

Comparison of Reported Methodological Characteristics Over Time

Selection of Studies

The two studies involved were investigations of the methodological characteristics within separate random samples of case-control studies investigating gene-disease associations. The first study, published in 2008 by Yesupriya et al[38] (including John P. A. Ioannidis (involved in the completion of this research)), analyzed the reporting characteristics of the study design, genotyping method, population stratification, analytic methodology and study inferences, analysis of multiple genetic variants, and analysis of interacting environmental factors from 315 randomly sampled studies published between 2001 and 2003 and contained within the Human Genome Epidemiology Network (HuGENet). In addition to this, 28 randomly sampled studies published in 2006 and contained within the Human Genome Epidemiology Network (HuGENet) were analyzed for their methodological characteristics and compared to the original set of studies.

The second study, performed in 2008 by AlJasir[39] (along with help from the current investigator), analyzed the effect of design and other basic study characteristics, population and methodological characteristics, outcome measure characteristics, and the reporting of Hardy-Weinberg equilibrium and statistical adjustment of covariates from 511 randomly sampled studies published in 2007 and contained within the Human Genome Epidemiology Network (HuGENet).

Statistical Analysis

Methodological characteristics that were identical and found to be extracted from the three random samples (genetic association studies published between 2001 and 2003 (Yesupriya et al[38]) vs.

2007 (Aljasir[39]), and genetic association studies published in 2006 (Yesupriya et al[38]) vs. 2007 (Aljasir[39])) in both studies were identified and compared to one another. Counts and percents were compared to one another by using the Chi-square test. All statistical analyses were performed using SAS 9.1 statistical software. An attempt was made to learn about the extraction methods used by both groups in order to identify whether or not these methods could be held responsible for the differences found between the results in the three samples of studies.

Genome-Wide Association Study Analysis

Selection of Studies

The sampling frame for genome-wide association studies included in this research was the National Human Genome Research Institute's (NHGRI) catalogue of published genome-wide association studies[4] (commenced on November 25, 2008), published in their respective scientific journals between 2005 and 2009 (420 potentially eligible journal articles for review). All genome-wide association studies presented included only those attempting to assay at least 100,000 single nucleotide polymorphisms (SNPs). Genome-wide association studies collected by the National Human Genome Research Institute's (NHGRI) catalog of published genome-wide association studies were identified through weekly PubMed literature searches, daily National Institutes of Health (NIH)-distributed compilations of news and media reports, and comparisons with an existing database of genome-wide association study literature (HuGE Navigator).[29,30,221]

It is important to understand that in analyzing this sample of genome-wide association studies, using the reported measure of uncertainty (p-value) as the outcome measure (see **Outcome Measure**, below) can be difficult when it comes to analyzing different studies with different diseases. The number of single nucleotide polymorphisms (SNPs) likely to have an effect on a genetic disease (e.g.,

Huntington's disease (one) vs. cancer (more than one)) makes it difficult to compare their relative measures of uncertainty (p-value). Therefore, in order to correct for this, this study used two samples of genome-wide association studies of two different categories of diseases (as defined by Manolio et al[31]) in order to prevent this from potentially manifesting itself in the results.

Manolio et al[31] identified seven different categories of diseases that are most prominent in the wealth of genome-wide association study literature (eye diseases, diabetes, cancer, gastrointestinal disorders, cardiovascular conditions and lipid metabolism, neuropsychiatric conditions, and autoimmune and infectious diseases). Based on the information contained within the study by Manolio et al,[31] in order to fulfill the objectives of this study, the current investigator chose genome-wide association studies in cancer (53 studies with relatively large sample sizes and relatively low odds ratios) and genome-wide association studies in neuropsychiatric conditions (74 studies with relatively small sample sizes and relatively high odds ratios). By choosing these two samples of studies (127 total studies[66-192]), the current investigator was able to analyze each group separately, and then to compare their results, in order to determine if a group of studies that demonstrate large sample sizes and low odds ratios show differences in methodological characteristics with a group of studies that demonstrate small sample sizes and high odds ratios.

Outcome Measure

Genome-wide association studies test the genome for single-nucleotide polymorphisms (SNPs) that may demonstrate evidence regarding susceptibility to a disease. These types of studies assume, prior to analysis, that certain regions of the genome may contain these single-nucleotide polymorphisms (SNPs). Since the objective of a genome-wide association study is to identify evidence for significant gene-disease associations, the primary outcome measure was the reported measure of uncertainty (p-value) for the genetic model considered in the primary analysis. In addition to this, since the objective of

the replication components embedded within a genome-wide association study is to identify if the association identified in the index components can be replicated (followed by an estimation of the magnitude of association given one exists), the secondary outcome measure was the reported odds ratio of the replication component. The primary and secondary outcome measures to be extracted, for both the index and replication components, were the most extreme reported measure of uncertainty (p-value) and odds ratio contained within the abstract of the study, respectively. If the study did not report a measure of uncertainty (p-value) or odds ratio within the abstract for either or both the index or replication components, then the entire study was reviewed to identify the most extreme reported measure of uncertainty (p-value) and odds ratio.

Language Restrictions

Only studies published in English were included in this research. Previous studies have shown that when a language restriction is included in medical research, the outcome measure is not altered by the introduction of any type of bias.[32,33] Additionally, some data have been interpreted as indicative of language biases and selective reporting in human genome epidemiology.[34] In view of these observations, and also due to time and language constraints, it was in the current investigator's best interest that the current research be performed with English studies only, to form a basis for future research in other languages.

Power Calculation

The calculation of power can be based on investigating previous studies designed to identify the effects of study characteristics in gene-disease association studies on the magnitude and direction of the effect estimate (e.g., odds ratio). Moher et al[36] have previously found that the reported median effect size was 0.5 (log odds ratio of -0.7 with a standard deviation of 1.13). In later investigative work,[37] they

considered a difference of 25% in the overall pooled log odds ratio as an indication of effect modification by study characteristic, assuming a random effects, two-sided t-test with a 95% confidence limit. Assuming these properties and a total genome-wide association study collection of 127 prospective journal articles for review, the current research had a power of 80.2% to detect a difference of 25% in the overall pooled log odds ratio as an indication of effect modification by study characteristic.

Pilot Study

Prior to the data extraction process a two-stage pilot study was performed in order to adapt the data extraction form, originally designed to extract data from case-control studies investigating gene-disease associations, into an instrument suitable for genome-wide association studies. The original data extraction form (Table 1) was developed by AlJasir[39] along with the help of the current investigator. In adapting this data extraction form for genome-wide association studies, it was necessary to take into account the fact that the number of genome-wide association studies is much less than the number of candidate gene studies, and also that there appears to be more scope for differences between genome-wide association studies in design, conduct and reporting. This pilot study also provided the two reviewers the opportunity to become familiar with the data extraction form, the data extraction process and to maximize inter-rater agreement.

Table 1: Data Extraction Form Developed for Candidate Gene Association Studies[39]

Study design and characteristics:

- Design of case-control study (classic or nested)
- Whether the case-control study is matched or not. (individually matched, frequency matched (one or several factors), not matched)
- Country of origin of research study. (Multi-centre was used if more than one country was involved).
- Journal (free text)
- Journal impact factor based on 2006
- Whether the association is claimed in the paper to be the first reported or a replication (first reported, replication, or it was mix or it was not reported)
- What was the number of genes included in the analysis?
- What was the number of genetic markers tested?
- Was the study investigating an association of a gene with cancer (Oncology) or a pre-cancer lesion? (Yes/ No)
- What was the disease category? (Allergy, andrology, breast, cardiology, dentistry, dermatology, ear, nose and throat (ENT), endocrinology, gastroenterology, gynecology and obstetric, hematology, hepatology, infectious diseases, metabolic disorder, nephrology, neurology, ophthalmology, orthopedic, prostate, psychiatric, respiratory, rheumatology, thyroid, urology and vascular).

Population and methodological characteristics:

- What was the final reported number of participants in the cases and control groups?
- What was the maximum number of potentially eligible participants that were invited to participate for cases and controls?
- Whether the control group is population-based from identified sampling frame, population-based without detail on the sampling frame, hospital and other specific group, control group was a mix of population and other specific group or cannot tell)
- What was the source of the data that was used in the study? (Primary collection and analysis of the data, building on a pre-existing data by collecting more information or blood sample, secondary analysis of a preexisting data or it wasn't clear).
- Was the process of handling the sample and genotyping described? (Yes for cases & controls, for cases only, for controls only, neither, can't tell)
- Was any method of quality assurance reported "e.g. blinding"? (Yes, No).
- Did the study report any result of calculating study power or sample size? (Yes, No).
- Did the study report the relationship between the control and cases groups (blood relation or relatives) for the specific reported outcome? (Yes, No, Unclear).

Outcomes analysis processes and Hardy Weinberg equilibrium (HWE):

- What was the measure of association reported in the abstract of the study (e.g. Odds ratio, 2X2 table, 2X3 table)?
- What was the specific reported number of participants in the cases and control groups associated with the extracted outcome?
- If an odds ratio was reported what was:
- The value reported in the abstract?
- The reported measure of uncertainty? Report the numeric measure of uncertainty that was used as well. (confidence limit, p-value, standard error and variance)
- If a 2X2 table was used, report:
- The numbers of cases with and without genetic exposure and the numbers of controls with and without genetic exposure?
- For a 2X3 table report the cases and control with "AA", "Aa" and "aa".
- Were the statistical methods for adjustment for the covariates specified? (Yes, No or it was not applicable).
- Did the study report any testing for the deviation from Hardy Weinberg equilibrium (HWE)? (Yes, No)
- If HWE was reported was the result? (All included genotypes were in equilibrium, some of the included

genotypes were in equilibrium (including the extracted outcome), some of the included genotypes were in equilibrium (not including the extracted outcome) and genotypes were not in equilibrium or it was not clear).

- Did the data allow for the calculation of HWE for the control, like presence of 2X3 table, p-value or inbreeding coefficient for the extracted outcome? (Yes, No)

Variables that were generated from the aforementioned extracted variables:

- Ratio of genetic markers per gene tested: This is equal to number of genetic markers tested per gene.
- Case / control ratio: This is the ratio of the final included number of cases to final number of included controls.
- Ratio of the excluded / included cases: This is equal to the difference between the maximum number of eligible cases and the final number of cases to the final included number of cases.
- Ratio of the excluded / included controls: This is equal to the difference between the maximum number of eligible controls and the final number of controls to the final included number of controls.
- Ratio of excluded / included sample size: This is equal to the difference between the maximum number of eligible sample size and the final number of sample size to the final included sample size.

Using the first draft of the data extraction form adapted to suit genome-wide association studies, the two reviewers applied the form to five journal articles[51-55] randomly selected from the National Human Genome Research Institute's (NHGRI) catalogue of published genome-wide association studies[4] and published between 2005 and 2009 (420 potentially eligible journal articles for review). The experience of using the form led to changes, resulting in the second draft.

Using the second draft of the data extraction form, the form was applied to ten journal articles,[56-65] again randomly selected from the NHGRI catalogue of published genome-wide association studies[4] and published between 2005 and 2009 (420 potentially eligible journal articles for review). In order to create the final data extraction form, a discussion regarding the results of the pilot study took place between the current investigator and the advisory committee, who have expertise in the fields of human genome epidemiology, and systematic reviews and meta-analyses. The results of both stages of the pilot study were carefully reviewed in order to identify how to optimize the final data extraction form, and the resulting modifications were agreed upon by all parties involved.

The final data extraction form (Table 2) consisted of seven sections: (1) General Characteristics, (2) Design of Index Component, (3) HWE Design/Usage of Index Component, (4) Study Results of Index Component, (5) Interpretation of Results of Index Component, (6) Replication Component (which includes sub-sections similar to sections (2) to (5)), and (7) Supplementary Material. From the production of the first draft of the data extraction form to the final data extraction form, no changes were made to sections (1), (3) and (5). Sections (2), (4), (6) and (7) were altered throughout the pilot study process, and the additions (presented in green) and deletions (presented in red) that were made are shown in Table 3.

Table 2: Final Data Extraction Form

GENERAL CHARACTERISTICS

1. Country of origin of study
2. Journal
3. Journal impact factor
4. Did the study investigate a gene with an association with cancer?
5. What disease was the study investigating?

DESIGN OF INDEX COMPONENT

6. Number of participants in case/control groups recruited
7. Was the control group:
    - Population-based with identified sampling frame,
    - Population-based with unidentified sampling frame,
    - Hospital or other specific group,
    - Mix of population-based and specific group,
    - Other/can't tell?
8. Was the data source:
    - Primary collection and analysis,
    - Building on pre-existing data through further collection,
    - Secondary analysis of pre-existing data?
9. Was the sample handling/genotyping process described?
    - If so, did the journal article reference methods from another journal article?
10. Was quality assurance reported?
11. Did the study report power/sample size calculations?
12. Were case/control groups related?
    - If so, were related cases/controls chosen directly or was cryptic relatedness identified?
13. Were cases/controls drawn from the same population?
14. Were extraction methods different for cases/controls?
15. How many SNPs were originally tested?

HWE DESIGN/USAGE OF INDEX COMPONENT

16. Was deviation from HWE tested?
17. If so, what was the result:
    - Equilibrium,
    - Non-equilibrium,
    - Can't tell?

STUDY RESULTS OF INDEX COMPONENT

18. What was the most extreme reported measure of uncertainty (e. g. p-value (include odds ratio))?
19. Was population stratification assessed?
20. Number of participants in case/control groups included in analysis
21. How many SNPs were included in the analysis?
22. What was the genome-wide threshold p-value?
23. How many SNPs crossed the genome-wide threshold p-value?
24. How many SNPs were included in the replication component?

INTERPRETATION OF RESULTS OF INDEX COMPONENT

25. Did the study detect/duplicate previously-known results?
26. Was evidence provided for a functional role of the polymorphism?

REPLICATION COMPONENT

27. Was a replication component performed?

28. Were the results replicated in an independent sample?
29. Were replication samples comparable to the original samples?

STUDY DESIGN

30. # of participants in case/control groups
31. Was the control group:
    - Population-based with identified sampling frame,
    - Population-based with unidentified sampling frame,
    - Hospital or other specific group,
    - Mix of population-based and specific group,
    - Other/can't tell?
32. Was the data source:
    - Primary collection and analysis,
    - Building of pre-existing data through further collection,
    - Secondary analysis of pre-existing data?
33. Was the sample handling/genotyping process described?
    - If so, did the journal article reference methods from another journal article?
34. Was quality assurance reported?
35. Did the study report power/sample size calculations?
36. Were case/control groups related?
    - If so, were related cases/controls chosen directly or was cryptic relatedness identified?
37. Were cases/controls drawn from the same population?
38. Were extraction methods different for cases/controls?

HWE DESIGN/USAGE

39. Was deviation from HWE tested?
40. If so, what was the result:
    - Equilibrium,
    - Non-equilibrium,
    - Can't tell?

STUDY RESULTS

41. What was the most extreme reported odds ratio?
42. Did the replication component confirm the results of the original study?

SUPPLEMENTARY MATERIAL

43. Did supplementary material accompany the journal article?
    - If so, how long was the reference material?

Table 3: List of Additions and Deletions Made During Development of the Data Extraction Form

Variable	*Reason for Addition/Deletion*
SECTION 2	
How many SNPs were originally tested?	To make a comparison between the SNPs included in the identification phase and the analysis phase of the index component, it is important to understand the number of SNPs originally identified for inclusion.
SECTION 4	
How were the results presented: ○ Allele/genotype frequencies, ○ Different genetic models, ○ Odds ratio, ○ Attributable risk, ○ p-value?	The process undertaken to identify the most suitable outcome measure rendered this question unnecessary.
What was the most extreme reported measure of uncertainty (e.g. p-value (include odds ratio))?	It was determined throughout the pilot study process that the most suitable outcome measure to extract from the index component was the most extreme reported measure of uncertainty (p-value).
Was a correction applied in defining the p-value?	The Bonferroni correction was used in each study to identify the genome-wide level of significance, rendering this question redundant.
Number of participants in case/control groups included in analysis.	To make a comparison between the participants originally recruited and those included in the analysis phase, it was important to understand the number of participants in case/controls groups included in the analysis.
How many SNPs were included in the analysis?	To make a comparison between the SNPs included in the identification phase and the analysis phase of the index component, it is important to understand the number of SNPs included in the analysis.
How many SNPs were included in the replication component?	To put the extracted information from both the index component and the replication component into context, it is important to understand the number of SNPs identified for replication.
SECTION 6	
What were the results? e.g. power, sample size, heterogeneity	Elaboration of the extraction process in the replication component rendered this question unnecessary.
SECTION 7	
Did supplementary material accompany the journal article? ○ If so, how long was the reference material?	Assessment of the reporting characteristics of each journal article required an in-depth look at how and where the information was being presented.

Data Extraction

Two reviewers partially independently extracted all information from the included studies using a data extraction form. The form to be used was adapted from a form previously developed by Aljasir[39] and his group of advisors performing an analysis involving case-control studies investigating gene-disease associations. As the data extraction form was originally designed to suit case-control studies investigating gene-disease associations, a more suitable data extraction form fit for genome-wide association studies was produced by adapting the original data extraction form in consultation with the same group of advisors.

The methodological features, identified as important in the interpretation of a genome-wide association study,[37] that were extracted from the genome-wide association studies included those related to study design, selection of study participants, genotyping and quality control, and analysis and presentation of results. Given that this research was mainly concerned with reporting quality in genome-wide association studies, extraction of information was done at the level of each journal article. Supplemental information was only accessed for data extraction purposes when the pre-determined method for identifying the outcome measures could not be achieved. An appropriate pilot study was performed with the extraction form prior to the actual data extraction process (see **Pilot Study**, above). Discrepancies in extracted information following the extraction process were partially resolved through consensus between the two reviewers.

Data Management

The extraction form was reproduced electronically and all included studies were collected for data extraction by using the form accordingly. Following data extraction and partial resolution of discrepancies in extracted information upon completion of the extraction process, the data were

transferred to SAS 9.1 statistical software and Biostat Comprehensive Meta-Analysis via Microsoft Excel 2007 in order to perform the statistical analysis.

Statistical Analysis

Descriptive statistics and graphical presentations of the data were produced in order to summarize the associations between the different variables and the outcome measure (reported measure of uncertainty (p-value) for the index components analysis and odds ratio for the replication components analysis). The two different disease groups were separated and analyzed individually in order to determine if the reporting practices differed between groups of investigators working on different types of diseases. Sensitivity analyses were performed in order to determine if the addition or removal of a group of studies with similar characteristics altered the combined effect of the studies in any way.

A meta-analysis was performed in order to identify the relative impact of the included variables in the studies selected, and the strength of the relationship between these variables and the outcome measure (reported measure of uncertainty (p-value) for the index components analysis and odds ratio for the replication components analysis). For the reported measure of uncertainty (p-value), Stouffer's Z-score was used as the summary measure in the index components of each study, while odds ratios were used as the summary measure in the replication components of each study. Stouffer's Z-score was chosen over Fisher's method as it can be used to summarize both one- and two-sided proportions, and it incorporates both the weight and direction of effect of each individual study.[232] Forest plots and bubble plots were produced, when possible, to demonstrate the results of the meta-analysis.

Funnel plots were produced in order to identify the potential for publication bias, and, if present, were analyzed using Egger's test[35] and Classic fail-safe N. Prior to analysis, it was hypothesized

that publication bias would more likely be an issue regarding the replication components rather than the index components, however both components of each study were analyzed for publication bias. In order to assess publication bias both within each group (index components vs. replication components) and between the groups (cancer vs. neuropsychiatric disorders), the results were grouped together and presented following the meta-analysis. The results of the test for publication bias were used to determine if the direction of effect found within each group of studies was related to study size.

As mentioned previously, all statistical analyses were performed using SAS 9.1 statistical software and Biostat Comprehensive Meta-Analysis via Microsoft Excel 2007.

Production of Guidelines

Following the completion of this research, it is the intent of the current investigator to consider how to improve the transparent reporting of methodological characteristics across all genome-wide association studies, with a view to proposing a possible adjunct to the STREGA (Strengthening the REporting of Genetic Association studies) Statement[5] regarding the reporting of genome-wide association studies.

SIGNIFICANCE OF THESIS

The purpose of this research was to perform a large-scale analysis of genome-wide association studies in order to produce evidence useful in guiding their reporting and synthesis. In addition to this, this research attempted to identify any obvious links between the methodological characteristics of case-control studies investigating gene-disease associations from two independent studies in two different time periods. It is the hope of the current investigator that this research will help to guide those performing future analyses of genome-wide association studies and identify the important methodological characteristics which influence the outcome of genome-wide association studies. It is also the hope of the current investigator that this research will help to guide those producing reporting guidelines specifically designed to assist in carrying out complete and transparent reporting of genome-wide association studies in the future.

RESULTS

Section 1: Comparison of Reported Methodological Characteristics Over Time

Information on seven methodological characteristics extracted from random samples of journal articles on genetic association studies published between 2001-2003 (Yesupriya et al[38]) and published in 2007 (Aljasir[39]) is compared in Table 4. In the more recent sample a higher proportion of authors stated that the journal article was the first report on the specific association (46.2% vs. 15.6%). The reported number of study participants increased. A higher proportion reported the statistical power of the study (74.4% vs. 12.7%). However, the proportion of studies for which clear information regarding the population from which cases and controls were drawn decreased (14.9% vs. 62.6%), as did the proportion that made an explicit statement about the use of unrelated study participants (1.2% vs. 17.8%). In a higher proportion of journal articles, it was reported that all genetic variants had been examined for Hardy-Weinberg equilibrium (83.2% vs. 47.9%). In the studies in which Hardy-Weinberg equilibrium was reported to have been examined, there was an increase in the proportion in which one or more polymorphisms were reported not to have been in equilibrium (20.9% vs. 6.6%).

Table 4: Comparison of the Methodological Characteristics of HuGE Journal Articles from Two Time Periods (2001-2003 vs. 2007)

Methodological Characteristic	*Published from 2001-2003*		*Published in 2007*		
	Count	*Percent*	*Count*	*Percent*	*P-Value*
Authors stated that this was the first study on the specific issue					<0.0001
Yes	49	15.6	236	46.2	
No	266	84.4	275	53.9	
Number of study participants					<0.0001
≤500	239	75.9	190	37.2	
501-999	47	14.9	139	27.2	
≥1000	29	9.2	182	35.6	
Reported the available power of the study					<0.0001
Yes	40	12.7	380	74.4	
No	275	87.3	131	25.6	
Reported clear information regarding the population from which cases and controls were drawn					<0.0001
Yes	142	62.6	76	14.9	
No	85	37.4	435	85.1	
Explicitly stated the use of unrelated study participants					<0.0001
Yes	56	17.8	6	1.2	
No	259	82.2	505	98.8	
Reported that all genetic variants were examined for Hardy-Weinberg equilibrium					<0.0001
Yes	151	47.9	425	83.2	
No	164	52.1	86	16.8	
If Hardy-Weinberg equilibrium was reported, did any polymorphism reportedly fail Hardy-Weinberg equilibrium					<0.0001
Yes	10	6.6	89	21.0	
No	141	93.4	336	79.0	

Information on four methodological characteristics extracted from random samples of journal articles on genetic association studies published in 2006 (Yesupriya et al[38]) and published in 2007 (AlJasir[39]) is compared in Table 5. Yesupriya et al[38] chose to analyze only four of the original seven methodological characteristics in the 2006 sample because these were addressed in less than 50% of the 2001-2003 journal articles. In the more recent sample, the reported number of study participants increased. A higher proportion reported the statistical power of the study (74.4% vs. 28.6%). However, the proportion that made an explicit statement about the use of unrelated study participants decreased (1.2% vs. 35.7%). In a higher proportion of journal articles, it was reported that all genetic variants had been examined for Hardy-Weinberg equilibrium (83.2% vs. 57.1%).

Table 5: Comparison of the Methodological Characteristics of HuGE Journal Articles from Two Time Periods (2006 vs. 2007)

Methodological Characteristic	*Published in 2006*		*Published in 2007*		
	Count	*Percent*	*Count*	*Percent*	*P-Value*
Number of study participants					
≤500	21	75.0	190	37.2	0.0003
501-999	4	14.3	139	27.2	
≥1000	3	10.7	182	35.6	
Reported the available power of the study					
Yes	8	28.6	380	74.4	<0.0001
No	20	71.4	131	25.6	
Explicitly stated the use of unrelated study participants					<0.0001
Yes	10	35.7	6	1.2	
No	18	64.3	505	98.8	
Reported that all genetic variants were examined for Hardy-Weinberg equilibrium					0.0007
Yes	16	57.1	425	83.2	
No	12	42.9	89	16.8	

Section 2: Review and Analysis of Reporting of Genome-Wide Association Studies

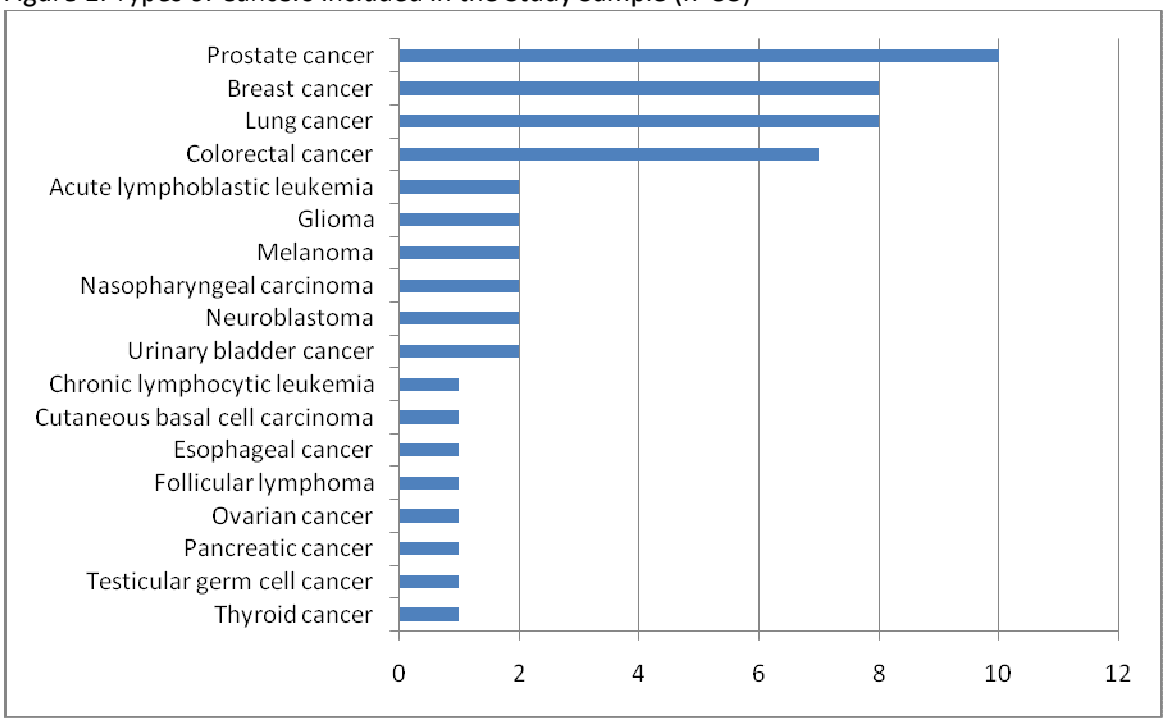
2.1 Review and Analysis of Reporting of Index Components of Cancer

Since the objective of the index components of a genome-wide association study is to discover gene-disease associations with a high degree of confidence that are not false-positive results, the primary outcome measure is the reported measure of uncertainty (p-value) for the genetic model considered in the primary analysis. If the index components did not report a measure of uncertainty (p-value) within the abstract, then the entire study was reviewed to identify the most extreme reported measure of uncertainty (p-value).

Description of Studies

The fifty-three genome-wide association studies of cancer that were selected for analysis[66-118] were published in 13 different journals whose median impact factor was 9.4 (range: 2.4 to 50.0). Multiple research centres in different countries were involved in 75.5% (n=40) of the studies, while 13.2% (n=7) were performed wholly in the USA, 9.4% (n=5) in the UK and 1.9% (n=1) in Japan. In the vast majority (86.8% (n=46)), supplementary material accompanied the journal article, with the median length of the supplementary material package being 14 pages (range: 4 to 211 pages). The studies reported on eighteen different types of cancer (Figure 1). The studies of acute lymphoblastic leukemia and neuroblastoma were done in children.

Figure 1: Types of Cancers Included in the Study Sample (n=53)



The index components of cancer typically recruited less than 5,000 study participants (67.9%, n=36; Table 6); the median number of study participants recruited was 3,403 (range: 60-37,553). In 58.5% (n=31) of studies, the control group was recruited from a hospital or another non-population-based specific group setting (e.g., blood donors), and the data source was built on pre-existing data through further collection 66.0% (n=35) of the time. Only one study failed to report the sample handling/genotyping process or methods of quality assurance (1.9%). The minority of studies, 13 of 53, reported on a power/sample size calculation relating to their study size (24.5%). In just under two-thirds of the studies (60.4%, n=32), case and control groups were explicitly reported as being unrelated to each other. In about half of the studies (52.8%, n=28) case and control groups were reported as having been chosen from the same source population. All of the studies reported using similar DNA extraction methods when collecting data from the case and control groups, however any specifics regarding quantity/quality of DNA were poorly reported overall (as determined through the pilot study process, it was difficult to identify any reasonable way to extract information regarding the specifics of the quantity/quality of DNA used in the index components). Finally, just over half (52.8%, n=28) of the studies investigated ≤500,000 SNPs, 41.5% (n=22) 500,000-999,999 SNPs and the remainder (5.7%, n=3) ≥1,000,000 SNPs. The median number of SNPs originally tested was 550,000 (range: 101,152 to 2,000,000).

Table 6: Design and Conduct of Index Components of Genome-Wide Association Studies in Cancer

Methodological Characteristic	*Count*	*Percent*
Number of study participants recruited:		
≤5000	36	67.9
5001-9999	7	13.2
≥10,000	10	18.9
Was the control group:		
Population-based with identified sampling frame,	12	22.6
Population-based with unidentified sampling frame,	0	0.0
Hospital or other non-population-based specific group (i.e. blood donors),	31	58.5
Mix of population-based and specific group,	10	18.9
Other/can't tell?	0	0.0
Was the data source:		
Primary collection and analysis,	14	26.4
Building on pre-existing data through further collection,	35	66.0
Secondary analysis of pre-existing data?	0	0.0
N/A	4	7.6
Was the sample handling/genotyping process described?		
Yes	52	98.1
No	1	1.9
Was quality assurance reported?		
Yes	52	98.1
No	1	1.9
Did the study report power/sample size calculations?		
Yes	13	24.5
No	40	75.5
Were case/control groups related?		
Yes (cryptic relationship)	14	26.4
Yes (direct relationship)	7	13.2
No	32	60.4
Were cases/controls drawn from the same population?		
Yes	25	47.2
No	28	52.8
Were DNA extraction methods different for cases/controls?		
Yes	0	0.0
No	53	100.0
How many SNPs were originally tested?		
≤500,000	28	52.8
500,000-999,999	22	41.5
≥1,000,000	3	5.7

Hardy-Weinberg equilibrium was assessed in the control groups in 50 of 53 (94.3%) of the index components of genome-wide association studies in cancer. Of these 50 studies, 8.0% found the SNPs of the control group satisfied equilibrium (n=4), 10.0% included the SNPs of the individuals in the control group that did not satisfy equilibrium (n=5) and 82.0% excluded the SNPs of the individuals in the control group that did not satisfy equilibrium (n=41).

As seen in Table 7, population stratification (by any means) was assessed in 67.9% (n=36) of the studies. As would be expected from the number of study participants recruited (Table 6), the data analyzed in the index components typically related to less than 5,000 study participants, while the median proportion of study participants whose data were included in analysis was 89.0% of those recruited. On average, 14.1% of SNPs originally selected for testing were excluded from analysis. The median number of SNPs analyzed was 462,866 (range: 60,275 to 2,000,000), compared with a median of 550,000 targeted by the arrays. Approximately half of the studies reported on the genome-wide threshold p-value (49.1%, n=26), and slightly less than half of the studies reported on the number of SNPs that crossed the genome-wide threshold p-value (43.4%, n=23). In the index components, previously-known results were reported to have been investigated and detected/duplicated in only 11.3% of the studies (n=6).

Table 7: Analysis of Index Components of Genome-Wide Association Studies in Cancer

Methodological Characteristic	*Count*	*Percent*
Was population stratification assessed?		
Yes	36	67.9
No	17	32.1
Number of study participants analyzed:		
≤5000	36	67.9
5001-9999	7	13.2
≥10,000	10	18.9
How many SNPs were included in the analysis?		
≤500,000	30	56.6
500,000-999,999	20	37.7
≥1,000,000	3	5.7
Was the genome-wide threshold p-value reported?		
Yes	26	49.1
No	27	50.9
Was the number of SNPs that crossed the genome-wide threshold p-value reported?		
Yes	23	43.4
No	30	56.6

Meta-Analysis

Out of the fifty-three genome-wide association studies in cancer that were chosen for analysis, forty were included in the meta-analysis. The list of excluded studies along with each study's reason for exclusion is presented in Table 8. In Spinola 2007[102] and Zanke 2007,[117] the most extreme reported measure of uncertainty (p-value) was not reported. Therefore, these two studies were not included in the meta-analysis.

Table 8: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Index Components of Genome-Wide Association Studies in Cancer

Study	*Journal*
Studies Excluded Because Direction of Effect Not Specified	
Duggan 2007[75]	J NATL CANCER INST
Easton 2007[76]	NATURE
Gold 2008[79]	PROC NATL ACAD SCI USA
Kibriya 2009[89]	BREAST CANCER RES TREAT
Murabito 2007[94]	BMC MED GENET
Murabito 2007[95]	BMC MED GENET
Skibola 2009[99]	NAT GENET
Thomas 2008[107]	NAT GENET
Tomlinson 2008[110]	NAT GENET
Tse 2009[112]	AM J HUM GENET
Wang 2008[113]	NAT GENET
Studies Excluded Because Most Extreme p-value Not Reported	
Spinola 2007[102]	CANCER LETT
Zanke 2007[117]	NAT GENET

The overall pooled regression coefficient produced from the meta-regression under a random effects model, Stouffer's Z-score and 95% confidence limit (a summary statistic used for the meta-analysis of proportions (e.g., p-value)), for the index components of genome-wide association studies in cancer was 0.073 (95% CI 0.058, 0.088). There was considerable heterogeneity between studies (Q 659.35 ($p < 0.001$); I^2 statistic 94.1).

The journal accounted for 10.66% of the total heterogeneity between studies, with a Q of 70.28 ($p = 1.29 \times 10^{-12}$). However, the only journals in which more than one study was included were *Cancer Research* (n=3) and *Nature Genetics* (n=31), for which the summary Stouffer's Z-scores were 0.054 (95% CI 0.071, 0.087) and 0.037 (95% CI 0.041, 0.044), respectively (Table 9 and Figure 2). Country of origin of study accounted for 4.64% of the total heterogeneity, with a Q of 30.62 ($p = 1.02 \times 10^{-6}$). Along with multi-centre studies, only those studies that originated in Japan (n=1) and the United States of America (n=5) demonstrated a significant summary Stouffer's Z-score, while those studies that originated in the United Kingdom (n=5) demonstrated an insignificant summary Stouffer's Z-score.

The type of disease accounted for 56.65% of total heterogeneity (Q 343.50 ($p = <1.00 \times 10^{-15}$)). Only the studies that were performed on acute lymphoblastic leukemia (n=2), breast cancer (n=4), colorectal cancer (n=5), lung cancer (n=6), prostate cancer (n=7) and urinary bladder cancer (n=2) showed significant heterogeneity. The type of disease that demonstrated the highest Stouffer's Z-score was nasopharyngeal cancer (n=1) (0.208 (95% CI 0.109, 0.304), while ovarian cancer (n=1) demonstrated the lowest Stouffer's Z-score (-0.071 (95% CI -0.087, -0.055).

Adjusting for the type of selection of the control group participants generated a between-study heterogeneity Q of 0.25 ($p = 0.882$). Adjusting for whether or not cases and controls were related, the between-study heterogeneity Q of 15.46 ($p = 4.40 \times 10^{-4}$) accounted for 2.34% of total heterogeneity. Studies that reported a cryptic relationship between the cases and controls demonstrated a summary

Stouffer's Z-score that bordered on insignificance. Reporting on whether or not cases and controls were drawn from the same population generated a between-study heterogeneity Q of 2.30 ($p = 0.129$).

The between-study heterogeneity when adjusting for type of data source generated a Q of 7.52 ($p = 0.023$), accounting for 2.30% of total heterogeneity. Studies that built their data on pre-existing data demonstrated a summary Stouffer's Z-score lower than the overall pooled Stouffer's Z-score, while studies that primarily collected their data demonstrated a summary Stouffer's Z-score higher than the overall pooled Stouffer's Z-score. Univariate meta-regression under a random effects model using unrestricted maximum likelihood, adjusting for reporting a power/sample size calculation, generated a between-study heterogeneity Q of 0.79 ($p = 0.373$).

Adjusting for reporting testing for deviation from Hardy-Weinberg equilibrium generated a between-study heterogeneity Q of 0.39 ($p = 0.532$). Among the studies that did report on Hardy-Weinberg equilibrium, the between-study heterogeneity generated a Q of 1.93 ($p = 0.382$). Reporting whether or not population stratification was assessed generated a between-study heterogeneity Q of <0.01 ($p = 0.975$).

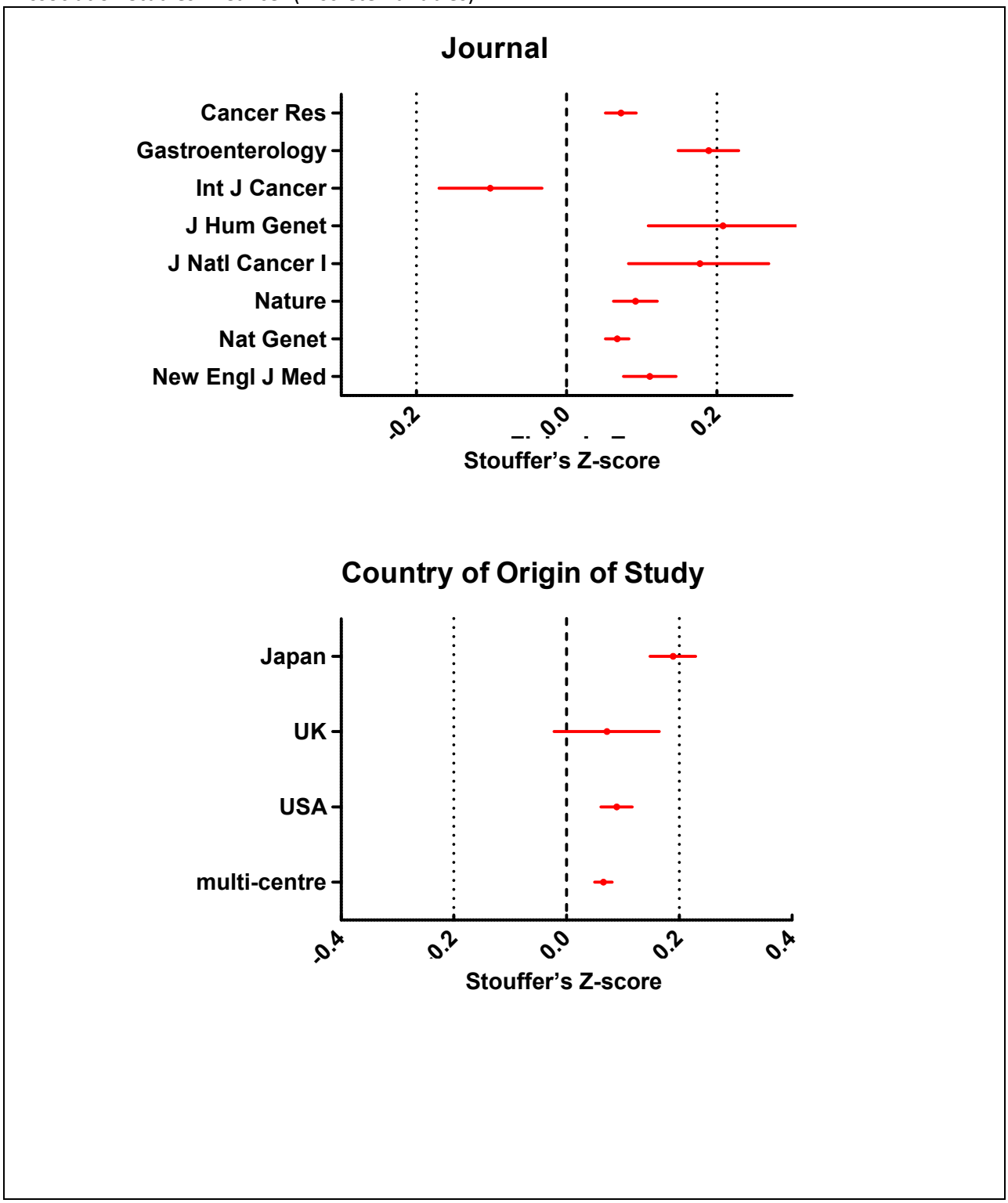
Adjusting for whether or not studies detected/duplicated previous results generated a between-study heterogeneity Q of 0.65 ($p = 0.422$). The between-study heterogeneity when adjusting for whether or not evidence was provided for a functional role of the polymorphism generated a Q of 0.16 ($p = 0.691$). Reporting whether or not supplementary material accompanied the journal article generated a between-study heterogeneity Q of 30.57 ($p = 3.23 \times 10^{-8}$), accounting for 5.40% of the total heterogeneity. Deviation from the overall Stouffer's Z-score was seen for both studies that included supplementary material and those that did not include supplementary material.

Table 9: Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Discrete Variables): 40 studies

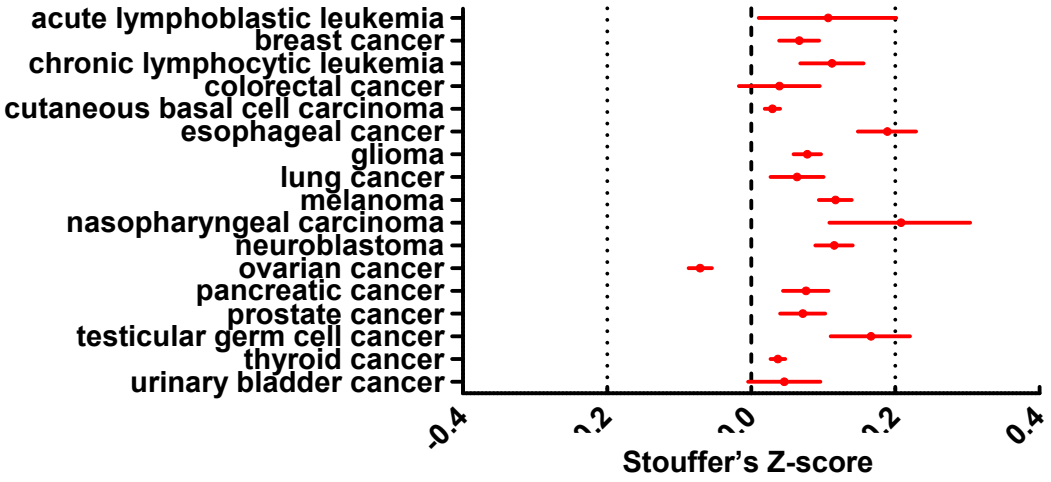
Methodological Characteristic	*Count*	*Stouffer's Z-score*	*95% CI*
Journal			
Cancer Research	3	0.072	(0.052, 0.093)
Gastroenterology	1	0.189	(0.148, 0.229)
International Journal of Cancer	1	-0.102	(-0.170, -0.033)
Journal of Human Genetics	1	0.208	(0.109, 0.304)
Journal of the National Cancer Institute	1	0.177	(0.082, 0.269)
Nature	1	0.092	(0.062, 0.121)
Nature Genetics	31	0.067	(0.052, 0.083)
The New England Journal of Medicine	1	0.111	(0.076, 0.146)
Country of Origin of Study			
Japan	1	0.189	(0.148, 0.229)
United Kingdom	5	0.072	(-0.022, 0.165)
United States of America	5	0.089	(0.061, 0.117)
Multi-centre	29	0.066	(0.050, 0.081)
What disease was the study investigating?			
Acute lymphoblastic leukemia	2	0.107	(0.010, 0.202)
Breast cancer	4	0.067	(0.039, 0.095)
Chronic lymphocytic leukemia	1	0.112	(0.068, 0.156)
Colorectal cancer	5	0.039	(-0.017, 0.095)
Cutaneous basal cell carcinoma	1	0.030	(0.019, 0.040)
Esophageal cancer	1	0.189	(0.148, 0.229)
Glioma	2	0.078	(0.059, 0.097)
Lung cancer	6	0.064	(0.027, 0.101)
Melanoma	2	0.117	(0.094, 0.140)
Nasopharyngeal carcinoma	1	0.208	(0.109, 0.304)
Neuroblastoma	2	0.115	(0.089, 0.141)
Ovarian cancer	1	-0.071	(-0.087, -0.054)
Pancreatic cancer	1	0.076	(0.044, 0.107)
Prostate cancer	7	0.072	(0.040, 0.103)
Testicular germ cell cancer	1	0.166	(0.111, 0.221)
Thyroid cancer	1	0.037	(0.027, 0.047)
Urinary bladder cancer	2	0.046	(-0.004, 0.096)
Was the control group:			
Population-based with identified sampling frame	7	0.072	(0.043, 0.102)
Hospital or other non-population-based specific group (i.e. blood donors)	25	0.077	(0.059, 0.095)
Mix of population-based and specific group	8	0.065	(0.018, 0.111)
Were case/control groups related?			
Yes (cryptic relationship)	12	0.100	(0.071, 0.129)
Yes (direct relationship)	7	0.039	(0.027, 0.052)
No	21	0.071	(0.039, 0.102)
Were cases/controls drawn from the same population?			
Yes	19	0.084	(0.066, 0.102)
No	21	0.063	(0.044, 0.083)
Was the data source:			
Primary collection and analysis	10	0.066	(0.049, 0.082)
Building on pre-existing data through further collection	27	0.107	(0.082, 0.133)
N/A	3	0.088	(0.046, 0.131)

Did the study report power/sample size calculations?			
Yes	9	0.058	(0.019, 0.098)
No	31	0.078	(0.062, 0.093)
Was deviation from HWE tested?			
Yes	33	0.076	(0.061, 0.091)
No	7	0.031	(-0.110, 0.171)
If so, what was the result?			
Equilibrium	3	0.149	(0.045, 0.250)
Non-equilibrium (controls included in further analyses)	4	0.081	(0.062, 0.099)
Non-equilibrium (controls excluded from further analyses)	26	0.076	(0.058, 0.093)
Was population stratification assessed?			
Yes	29	0.043	(0.039, 0.047)
No	11	0.045	(0.039, 0.052)
Did the study detect/duplicate previously known results?			
Yes	6	0.066	(0.025, 0.107)
No	34	0.084	(0.066, 0.103)
Was evidence provided for a functional role of the polymorphism?			
Yes	14	0.085	(0.053, 0.116)
No	26	0.077	(0.060, 0.094)
Did supplementary material accompany the journal article?			
Yes	37	0.076	(0.060, 0.092)
No	3	0.192	(0.154, 0.229)

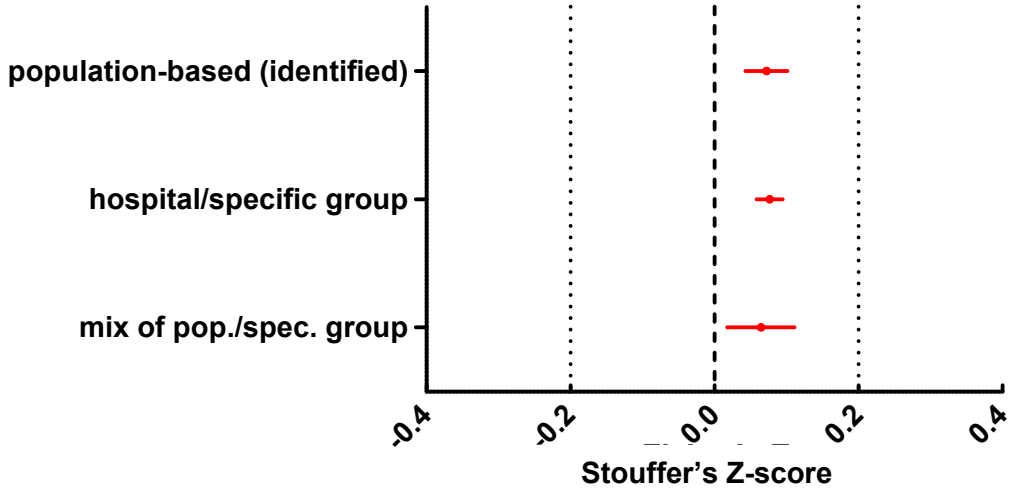
Figure 2: Forest Plots of Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Discrete Variables)



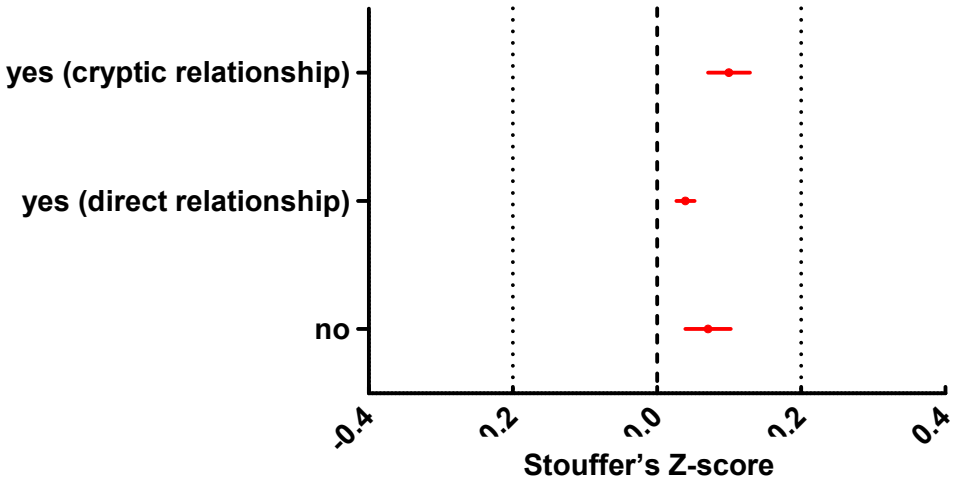
What disease was the study investigating?



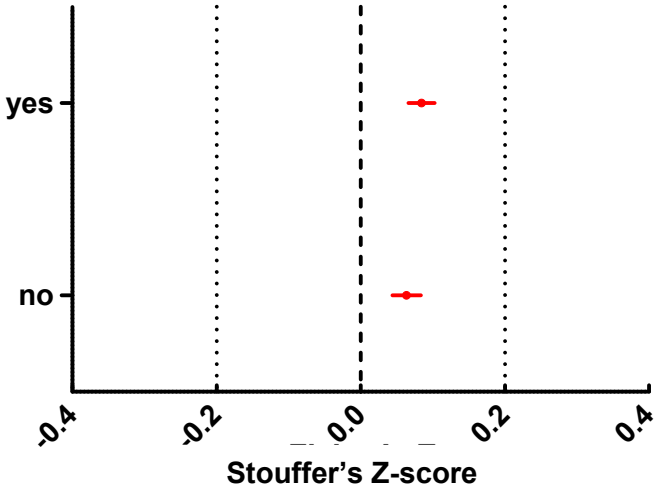
Was the control group:



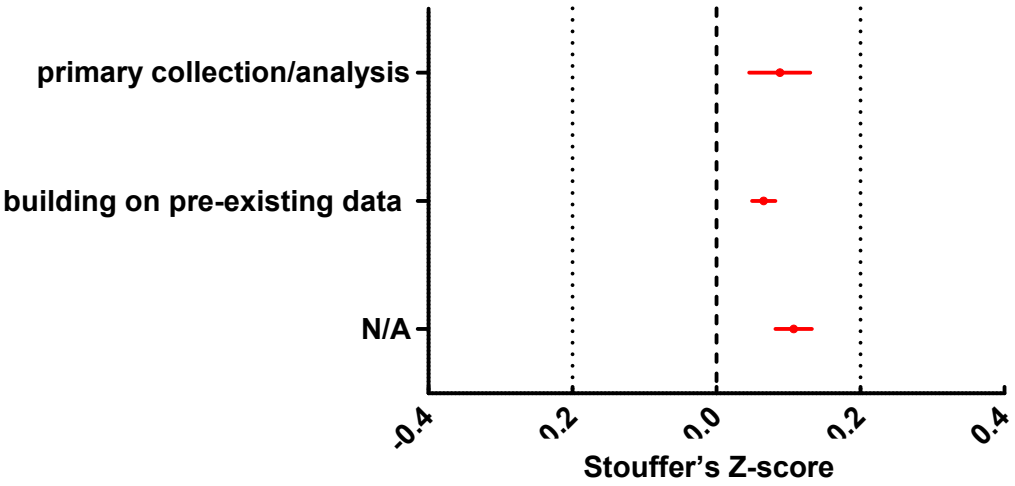
Were case/control groups related?



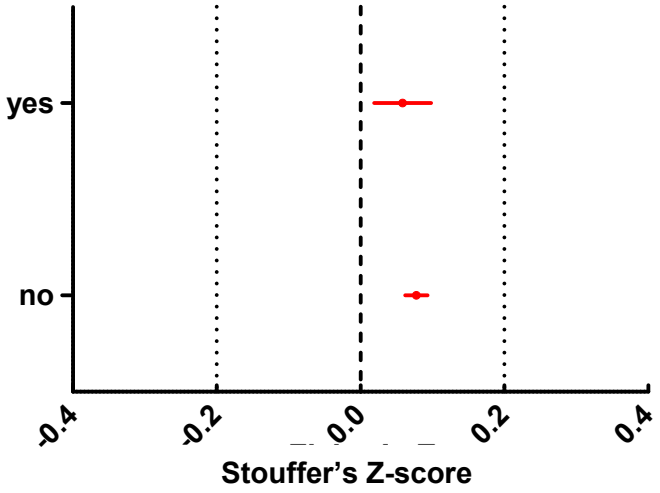
Were cases/controls drawn from the same population?



Was the data source:

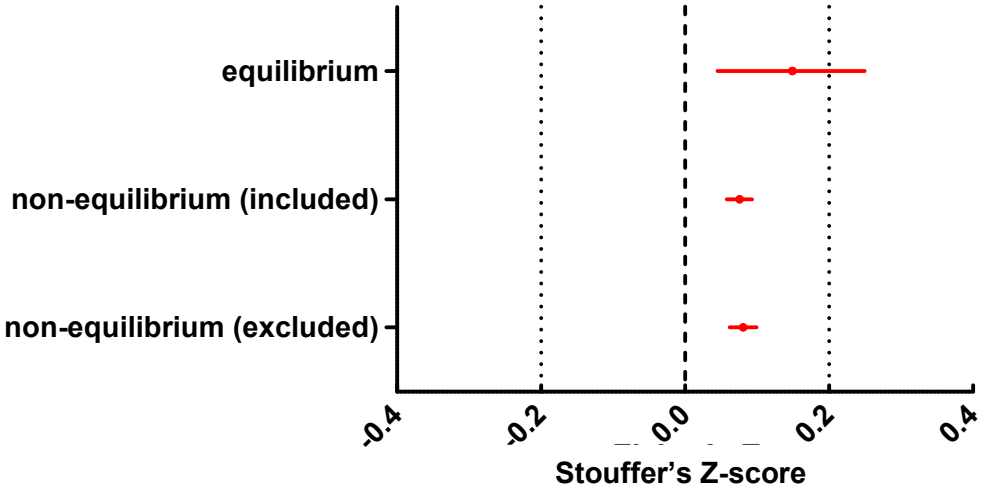


Did the study report power/sample size calculations?

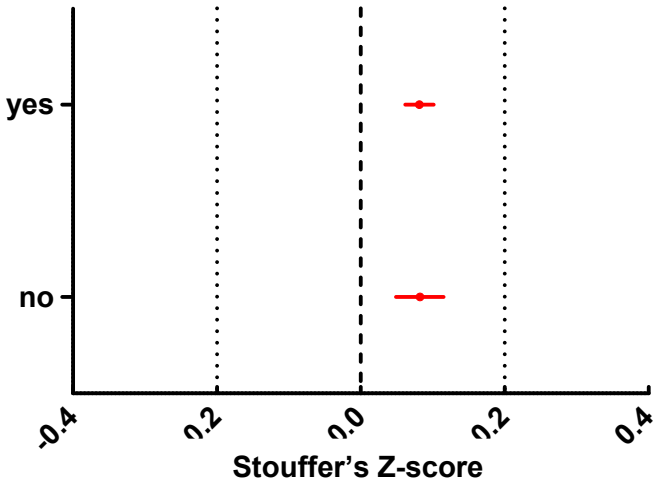


Was deviation from HWE tested?

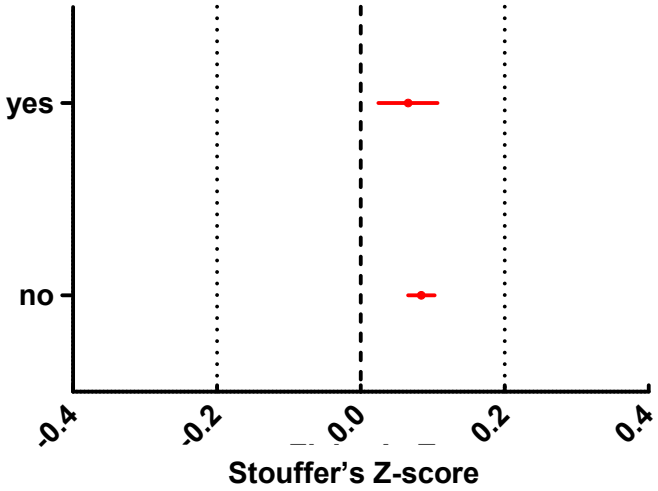
If so, what was the result?



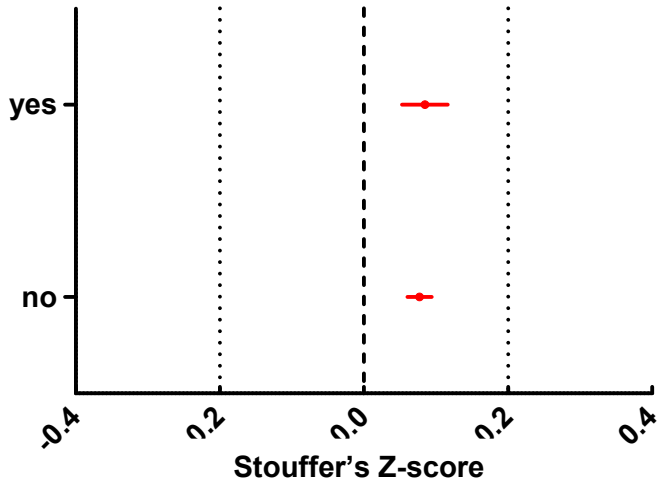
Was population stratification assessed?



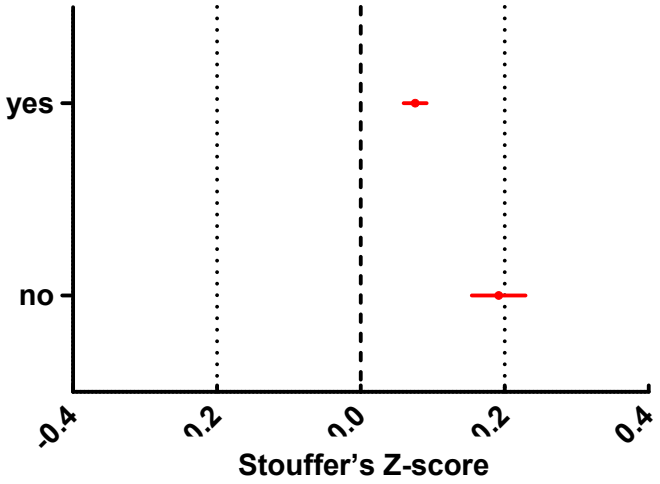
Did the study detect/duplicate previously known results?



Was evidence provided for a functional role of the polymorphism?



Did supplementary material accompany the journal article?



Univariate meta-regression under a random effects model using unrestricted maximum likelihood, adjusting for journal impact factor, generated a between-study heterogeneity Q of 0.13 (p = 0.721) (Table 10, Figure 3).

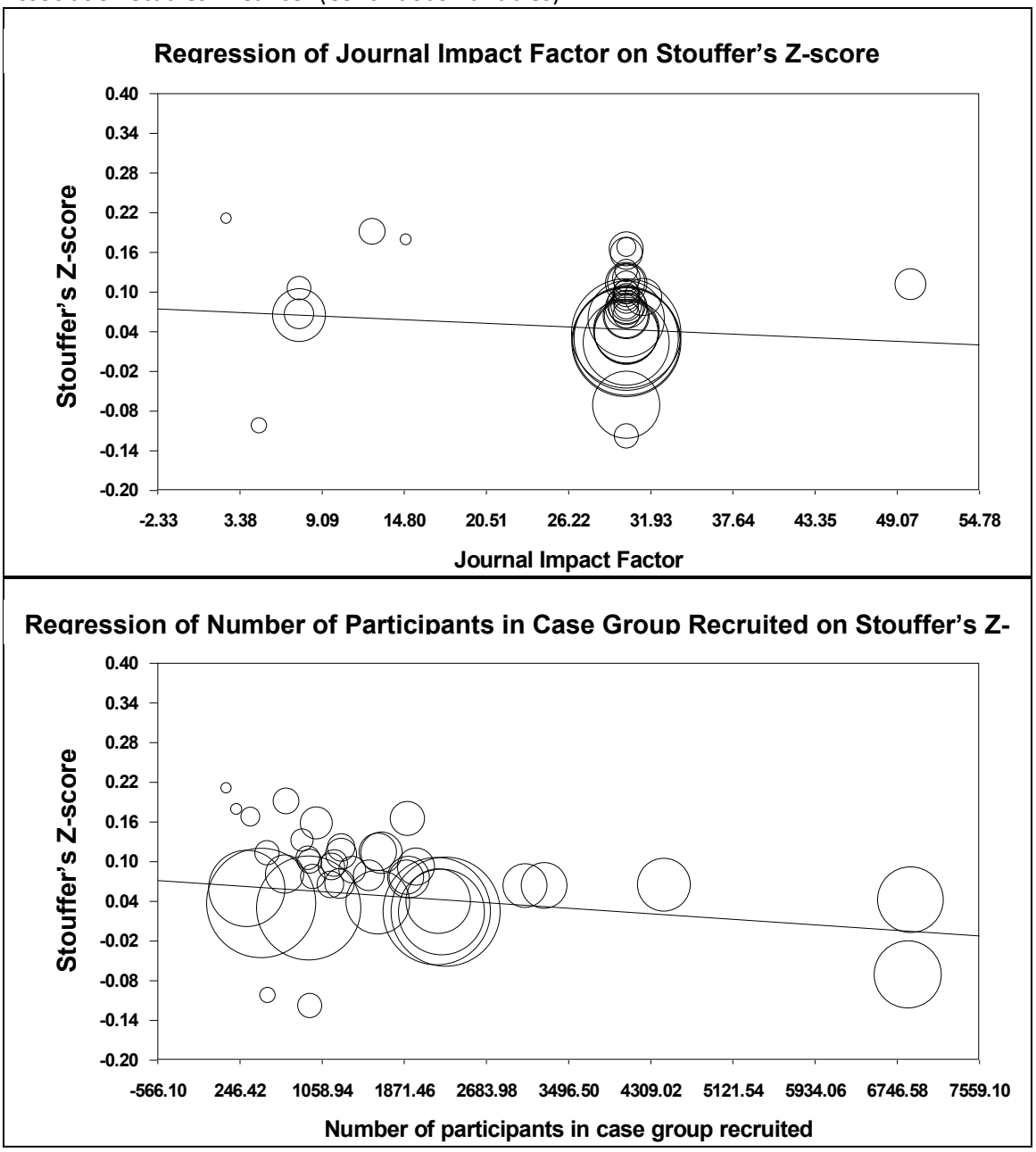
Adjusting for the number of participants recruited for the case group, the between-study heterogeneity Q of 6.73 (p = 0.009) accounted for 1.02% of total heterogeneity. Reported Stouffer's Z-scores tended to decrease as the number of participants in the case group recruited increased. Reporting the number of participants in the case group analyzed generated a between-study heterogeneity Q of 6.16 (p = 0.013), 0.93% of the total heterogeneity. Reported Stouffer's Z-scores tended to decrease as the number of participants in the control group analyzed increased. Adjusting for the number of participants recruited for the control group, the between-study heterogeneity Q of 6.19 (p = 0.013) accounted for 0.94% of total heterogeneity. Reported Stouffer's Z-scores tended to decrease as the number of participants in the control group recruited increased. Reporting the number of participants in the control group analyzed generated a between-study heterogeneity Q of 6.11 (p = 0.013), 0.93% of the total heterogeneity. Reported Stouffer's Z-scores tended to decrease as the number of participants in the control group analyzed increased.

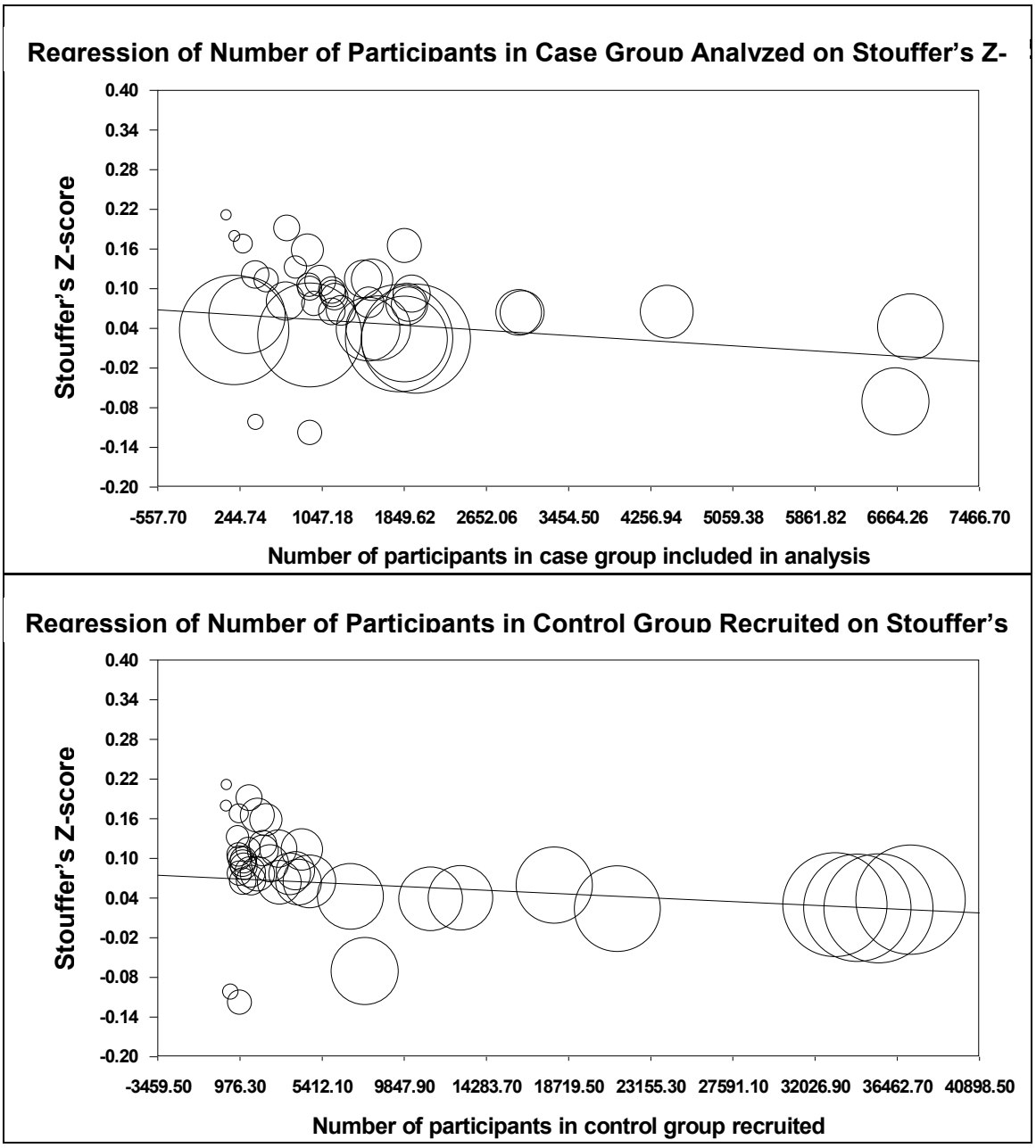
Adjusting for the number of SNPs originally tested, the between-study heterogeneity Q was 1.39 (p = 0.239). Reporting the number of SNPs included in the analysis generated a between-study heterogeneity Q of 0.25 (p = 0.618). Adjusting for the length of reference material included (given that reference material was included), the between-study heterogeneity Q was 0.01 (p = 0.924).

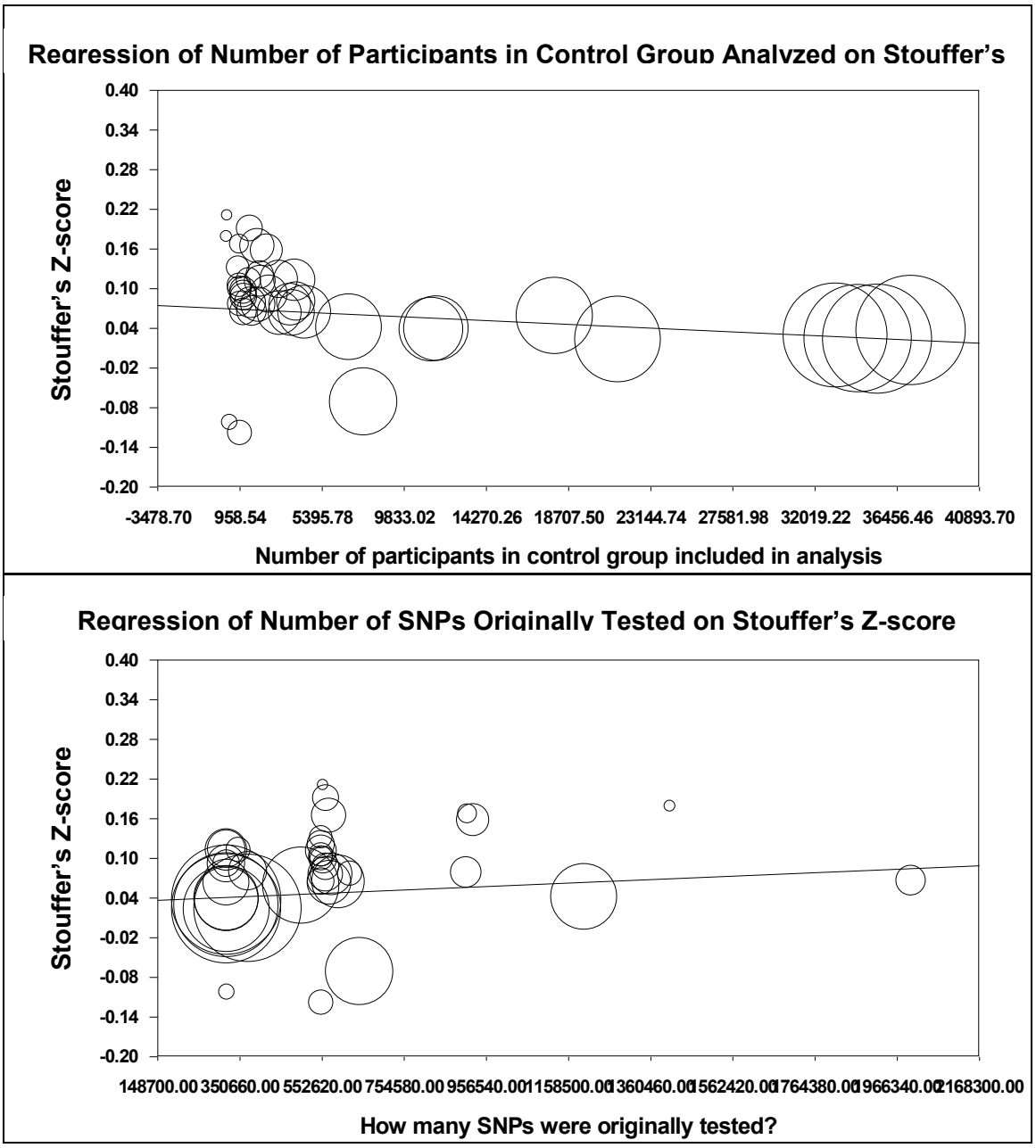
Table 10: Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Continuous Variables)

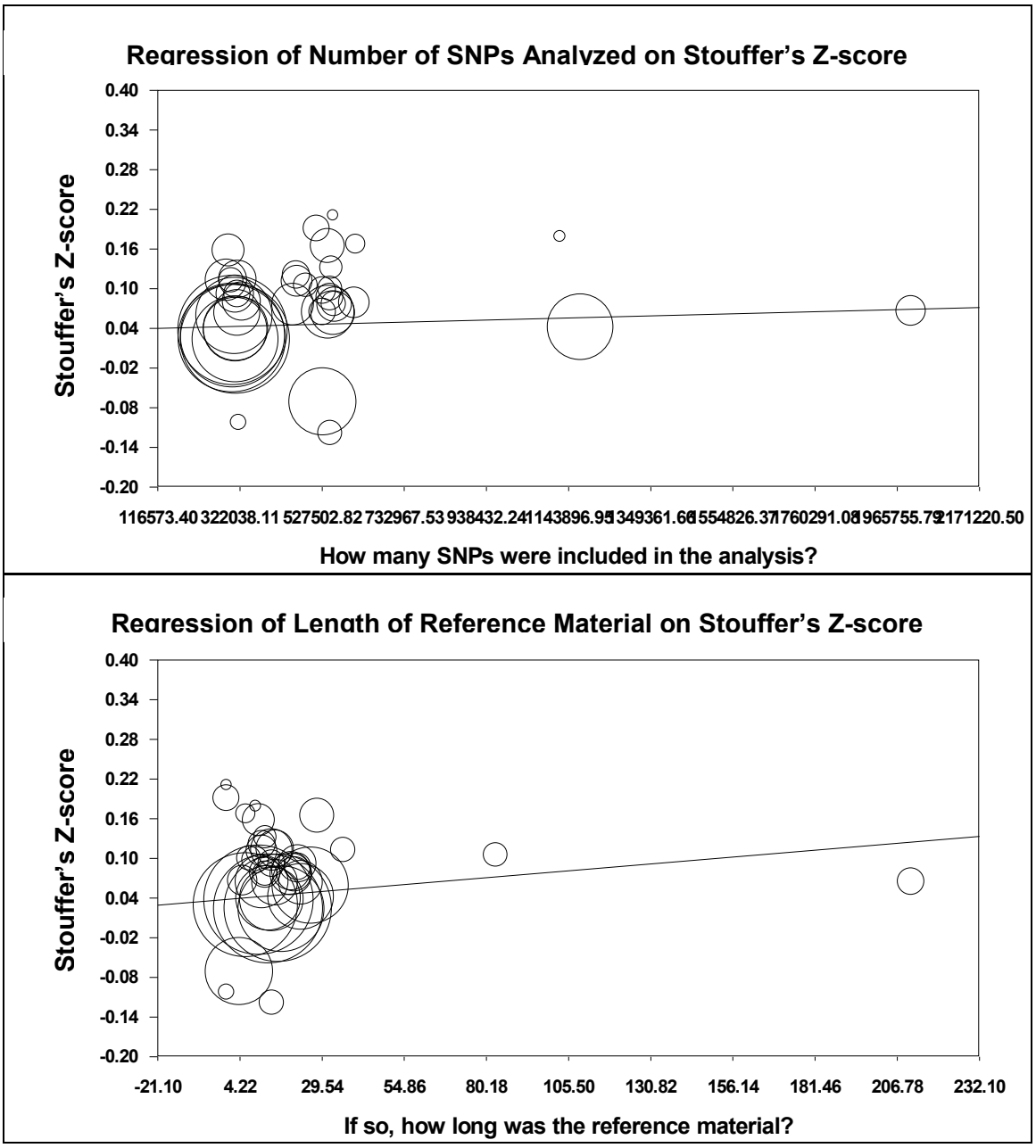
Methodological Characteristic	*Slope*	*95% CI*	*Q-value (p-value)*
Journal Impact Factor	-0.00095	(-0.00159, -0.00032)	0.13 (0.721)
Number of study participants in case group recruited	-0.00001	(-0.00001, -0.00001)	6.73 (0.009)
Number of study participants in case group analyzed	-0.00001	(-0.00001, -0.00001)	6.16 (0.013)
Number of study participants in control group recruited	<-0.00001	(<-0.00001, <-0.00001)	6.19 (0.013)
Number of study participants in control group analyzed	<-0.00001	(<-0.00001, <-0.00001)	6.11 (0.013)
How many SNPs were originally tested?	<0.00001	(<0.00001, <0.00001)	1.39 (0.239)
How many SNPs were included in the analysis?	<0.00001	(<0.00001, <0.00001)	0.25 (0.618)
If reference material was included, how long was it?	0.00041	(0.00022, 0.00060)	0.01 (0.924)

Figure 3: Scatter Plots of Results of Meta-Regression Analysis of Index Components of Genome-Wide Association Studies in Cancer (Continuous Variables)




Regression of Number of Participants in Case Group Analyzed on Stouffer's Z-
Stouffer's Z-score
0.40
0.34
0.28
0.22
0.16
0.10
0.04
-0.02
-0.08
-0.14
-0.20
-557.70
244.74
1047.18
1849.62
2652.06
3454.50
4256.94
5059.38
5861.82
6664.26
7466.70
Number of participants in case group included in analysis
Regression of Number of Participants in Control Group Recruited on Stouffer's
Stouffer's Z-score
0.40
0.34
0.28
0.22
0.16
0.10
0.04
-0.02
-0.08
-0.14
-0.20
-3459.50
976.30
5412.10
9847.90
14283.70
18719.50
23155.30
27591.10
32026.90
36462.70
40898.50
Number of participants in control group recruited


Regression of Number of Participants in Control Group Analyzed on Stouffer's
Stouffer's Z-score
0.40
0.34
0.28
0.22
0.16
0.10
0.04
-0.02
-0.08
-0.14
-0.20
-3478.70
958.54
5395.78
9833.02
14270.26
18707.50
23144.74
27581.98
32019.22
36456.46
40893.70
Number of participants in control group included in analysis
Regression of Number of SNPs Originally Tested on Stouffer's Z-score
Stouffer's Z-score
0.40
0.34
0.28
0.22
0.16
0.10
0.04
-0.02
-0.08
-0.14
-0.20
148700.00
350660.00
552620.00
754580.00
956540.00
1158500.00
1360460.00
1562420.00
1764380.00
1966340.00
2168300.00
How many SNPs were originally tested?


Regression of Number of SNPs Analyzed on Stouffer's Z-score
Stouffer's Z-score
0.40
0.34
0.28
0.22
0.16
0.10
0.04
-0.02
-0.08
-0.14
-0.20
116573.40
322038.11
527502.82
732967.53
938432.24
1143896.95
1349361.66
1554826.37
1760291.08
1965755.79
2171220.50
How many SNPs were included in the analysis?
Regression of Length of Reference Material on Stouffer's Z-score
Stouffer's Z-score
0.40
0.34
0.28
0.22
0.16
0.10
0.04
-0.02
-0.08
-0.14
-0.20
-21.10
4.22
29.54
54.86
80.18
105.50
130.82
156.14
181.46
206.78
232.10
If so, how long was the reference material?

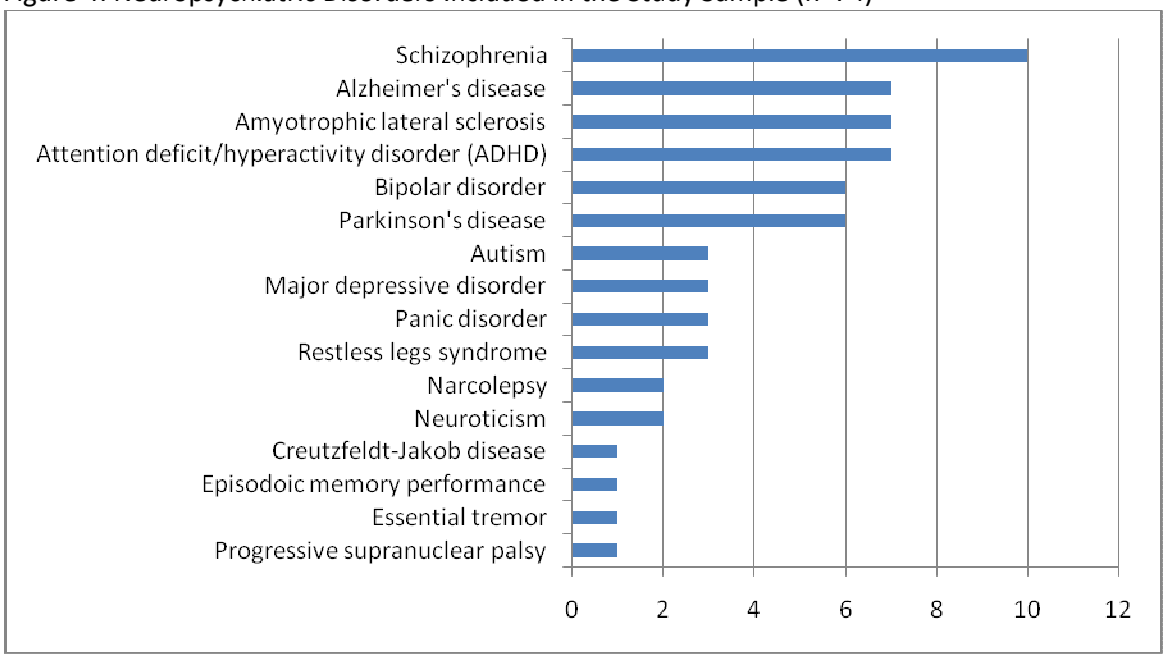
2.2 Review and Analysis of Reporting of Index Components of Neuropsychiatric Disorders

Since the objective of the index components of a genome-wide association study is to discover gene-disease associations with a high degree of confidence that are not false-positive results, the primary outcome measure is the reported measure of uncertainty (p-value) for the genetic model considered in the primary analysis. If the index components did not report a measure of uncertainty (p-value) within the abstract, then the entire study was reviewed to identify the most extreme reported measure of uncertainty (p-value).

Descriptive Statistics

The seventy-four genome-wide association studies of neuropsychiatric disorders that were selected for analysis[119-192] were published in 24 different journals whose median impact factor was 7.2 (range: 2.2 to 50.0). Multiple research centres in different countries were involved in 74.3% (n=55) of the studies, while 20.3% (n=15) were performed wholly in the USA, 4.1% (n=3) in Japan and 1.4% (n=1) in the UK. In the majority (67.6% (n=50)), supplementary material accompanied the journal article, with the median length of the supplementary material package being 17 pages (range: 2 to 443 pages). The studies reported on sixteen different neuropsychiatric disorders (Figure 5).

Figure 4: Neuropsychiatric Disorders Included in the Study Sample (n=74)



The index components of neuropsychiatric diseases typically recruited less than 5,000 study participants (82.4%, n=61; Table 11); the median number of study participants recruited was 1,965.5 (range: 71-18,451). In 62.2% (n=46) of studies, the control group was recruited from a hospital or another non-population-based specific group setting (e.g., blood donors), and the data source was built on pre-existing data through further collection 45.9% (n=34) of the time. Only two studies failed to report the sample handling/genotyping process or methods of quality assurance (2.7%). The minority of studies, 12 of 74, reported on quality assurance (16.2%) and only 28 studies reported on a power/sample size calculation relating to their study size (37.8%). In two-thirds of the studies (66.2%, n=49), case and control groups were explicitly reported as being unrelated to each other. In just under two-thirds of the studies (62.2%, n=46) case and control groups were reported as having been not chosen from the same source population. All of the studies reported using similar DNA extraction methods when collecting data from the case and control groups, however any specifics regarding quantity/quality of DNA were poorly reported overall (as determined through the pilot study process, it was difficult to identify any reasonable way to extract information regarding the specifics of the quantity/quality of DNA used in the index components). Finally, 20.3% (n=15) of the studies investigated ≤500,000 SNPs, 70.3% (n=52) 500,000-999,999 SNPs and the remainder (9.4%, n=7) ≥1,000,000 SNPs. The median number of SNPs originally tested was 550,000 (range: 100,000 to 1,631,003).

Table 11: Design and Conduct of Index Components of Genome-Wide Association Studies in Neuropsychiatric Disorders

Methodological Characteristic	*Count*	*Percent*
Number of study participants recruited:		
≤5000	61	82.4
5001-9999	5	6.8
≥10,000	8	10.8
Was the control group:		
Population-based with identified sampling frame,	16	21.6
Population-based with unidentified sampling frame,	0	0.0
Hospital or other non-population-based specific group (i.e. blood donors),	46	62.2
Mix of population-based and specific group,	5	6.8
Other/can't tell?	7	9.5
Was the data source:		
Primary collection and analysis,	27	36.5
Building on pre-existing data through further collection,	0	45.9
Secondary analysis of pre-existing data?	34	0.00
N/A	13	17.6
Was the sample handling/genotyping process described?		
Yes	72	97.3
No	2	2.7
Was quality assurance reported?		
Yes	62	83.8
No	12	16.2
Did the study report power/sample size calculations?		
Yes	28	37.8
No	46	62.2
Were case/control groups related?		
Yes (cryptic relationship)	19	25.7
Yes (direct relationship)	6	8.1
No	49	66.2
Were cases/controls drawn from the same population?		
Yes	28	37.8
No	46	62.2
Were DNA extraction methods different for cases/controls?		
Yes	0	0.0
No	74	100.0
How many SNPs were originally tested?		
≤500,000	15	20.3
500,000-999,999	52	70.3
≥1,000,000	7	9.4

Hardy-Weinberg equilibrium was assessed in the control groups in 57 of 74 (77.0%) of the index components of genome-wide association studies in neuropsychiatric disorders. Of these 57 studies, 7.0% found the SNPs of the control group satisfied equilibrium (n=4), 10.5% included the SNPs of the individuals in the control group that did not satisfy equilibrium (n=6) and 82.5% excluded the SNPs of the individuals in the control group that did not satisfy equilibrium (n=47).

As seen in Table 12, population stratification (by any means) was assessed in 67.6% (n=50) of the studies. As would be expected from the number of subjects recruited (Table 11), the data analyzed in the index components typically related to less than 5,000 study participants, while the median proportion of study participants whose data were included in analysis was 95.9% of those recruited. On average, 15.3% of SNPs originally selected for testing were excluded from analysis. The median number of SNPs analyzed was 437,127 (range: 100,000 to 1,564,797), compared with a median of 550,000 targeted by the arrays. The majority of studies reported on both the genome-wide threshold p-value (70.3%, n=52), and the number of SNPs that crossed the genome-wide threshold p-value (70.3%, n=52). In the index components, previously-known results were reported to have been investigated and detected/duplicated in only 1.4% of the studies (n=1).

Table 12: Analysis of Index Components of Genome-Wide Association Studies in Neuropsychiatric Disorders

Methodological Characteristic	*Count*	*Percent*
Was population stratification assessed?		
Yes	50	67.6
No	24	32.4
Number of study participants analyzed:		
≤5000	63	85.1
5001-9999	4	5.41
≥10,000	7	9.46
How many SNPs were included in the analysis?		
≤500,000	46	62.2
500,000-999,999	22	29.7
≥1,000,000	6	8.1
Was the genome-wide threshold p-value reported?		
Yes	52	70.3
No	22	29.7
Was the number of SNPs that crossed the genome-wide threshold p-value reported?		
Yes	52	70.3
No	22	29.7

Meta-Analysis

Out of the seventy-four genome-wide association studies in neuropsychiatric disorders that were chosen for analysis, thirty-six were included in the meta-analysis. The list of excluded studies along with each study's reason for exclusion is presented in Table 13. In Dunckley 2007[130] and Zhang 2007,[192] the most extreme reported measure of uncertainty (p-value) was not reported. Therefore, these two studies were not included in the meta-analysis.

Table 13: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders

Study	*Journal*
Studies Excluded Because Direction of Effect Not Specified	
Abraham 2008[119]	BMC MED GENOM
Anney 2008[120]	AM J MED GENET B
Baum 2008[121]	MOL PSYCHIATR
Beecham 2009[122]	AM J HUM GENET
Bertram 2008[123]	AM J HUM GENET
Blauw 2008[124]	LANCET NEUROL
Chio 2009[126]	HUM MOL GENET
Coon 2007[127]	J CLIN PSYCHIAT
Cronin 2008[128]	HUM MOL GENET
Ferreira 2008[131]	NAT GENET
Feulner 2009[132]	MOL PSYCHIATR
Hattori 2009[136]	AM J MED GENET B
Heinzen 2009[137]	J ALZHEIMERS DIS
Huentelmann 2007[138]	HUM MOL GENET
Kirov 2009[140]	MOL PSYCHIATR
Landers 2009[142]	PROC NATL ACAD SCI USA
Lasky-Su 2008[143]	AM J MED GENET B
Lasky-Su 2008[144]	AM J MED GENET B
Latourelle 2009[145]	BMC MED GENET
Lesch 2008[147]	J NEURAL TRANSM
Li 2008[148]	ARCH NEUROL
Ma 2009[149]	ANN HUM GENET
Mead 2008[151]	LANCET NEUROL
Mick 2008[153]	AM J MED GENET B
Muglia 2008[155]	MOL PSYCHIATR
Need 2009[157]	PLOS GENET
Otowa 2009[159]	J HUM GENET
Poduslo 2009[161]	AM J MED GENET B
Potkin 2009[162]	PLOS ONE
Shi 2009[169]	NATURE
Shifman 2008[170]	MOL PSYCHIATR
Shyn 2009[172]	MOL PSYCHIATR
Sonuga-Barke 2008[176]	AM J MED GENET B
Sonuga-Barke 2008[177]	AM J MED GENET B
van den Oord 2008[183]	ARCH GEN PSYCHIAT
Winkelmann 2007[191]	NAT GENET
Studies Excluded Because Most Extreme p-value Not Reported	
Dunckley 2007[130]	NEW ENGL J MED
Zhang 2009[192]	MOL PSYCHIATR

The overall pooled regression coefficient produced from the meta-regression under a random effects model, Stouffer's Z-score and 95% confidence limit (a summary statistic used for the meta-analysis of proportions (e.g., p-value)), for the index components of genome-wide association studies in neuropsychiatric disorders was 0.072 (95% CI 0.046, 0.099). There was considerable heterogeneity between studies (Q 983.05 ($p < 0.001$); I^2 statistic 96.4).

The journal accounted for 6.69% of the total heterogeneity between studies, with a Q of 65.80 ($p = 4.72 \times 10^{-9}$). However, the only journals in which more than three studies were included were *Molecular Psychiatry* (n=5), *Nature* (n=5), and *Nature Genetics* (n=12), for which the summary Stouffer's Z-scores were 0.015 (95% CI -0.003, 0.033), 0.101 (95% CI 0.093, 0.108) and 0.171 (95% CI 0.120, 0.221) respectively. Country of origin of study demonstrated a Q of 6.70 ($p = 0.082$).

The type of disease accounted for 16.61% of total heterogeneity (Q 163.25 ($p = <1.00 \times 10^{-15}$)). Only the studies that were performed on Alzheimer's disease (n=4), amyotrophic lateral sclerosis (n=6), autism (n=2), bipolar disorder (n=4), Parkinson's disease (n=5) and schizophrenia (n=7) showed significant heterogeneity. The type of disease that demonstrated the highest Stouffer's Z-score was narcolepsy (n=2) (0.165 (95% CI 0.126, 0.203)), while attention deficit/hyperactivity disorder (n=6) demonstrated the lowest Stouffer's Z-score (-0.086 (95% CI -0.123, -0.048)).

Adjusting for the type of selection of the control group participants generated a between-study heterogeneity Q of 3.51 ($p = 0.320$). Adjusting for whether or not cases and controls were related, the between-study heterogeneity Q was 0.16 ($p = 0.925$). Reporting on whether or not cases and controls were drawn from the same population generated a between-study heterogeneity Q of 0.67 ($p = 0.413$).

The between-study heterogeneity when adjusting for type of data source generated a Q of 1.23 ($p = 0.539$). Univariate meta-regression under a random effects model using unrestricted maximum

likelihood, adjusting for reporting a power/sample size calculation, generated a between-study heterogeneity Q of 5.16 ($p = 0.023$), and showed that between-study heterogeneity accounted for 0.53% of total heterogeneity. Studies that reported a power/sample size calculation demonstrated a summary Stouffer's Z-score that was dissimilar from the overall pooled Stouffer's Z-score.

Adjusting for reporting testing for deviation from Hardy-Weinberg equilibrium generated a between-study heterogeneity Q of 0.03 ($p = 0.960$). Among the studies that did report on Hardy-Weinberg equilibrium, the between-study heterogeneity generated a Q of 2.75 ($p = 0.097$). Reporting whether or not population stratification was assessed generated a between-study heterogeneity Q of <0.01 ($p = 0.984$).

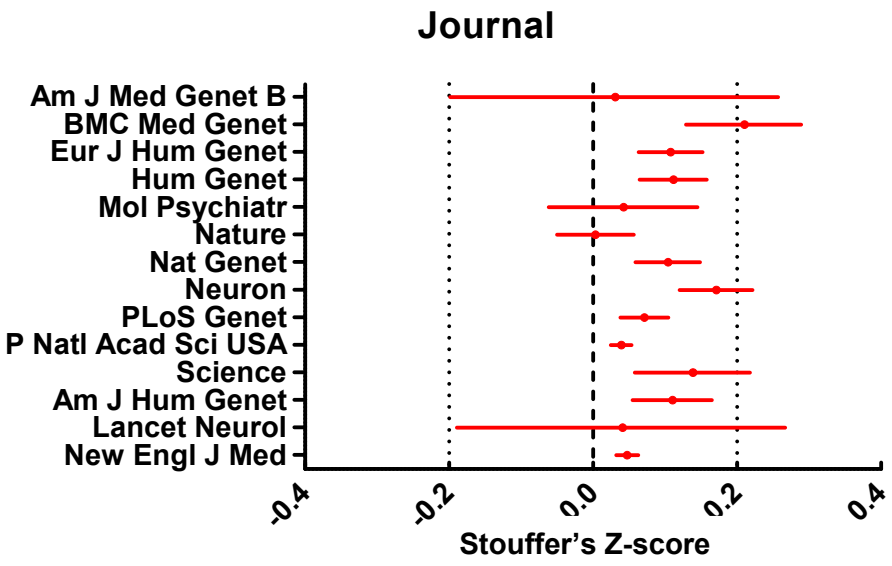
Adjusting for whether or not studies detected/duplicated previous results generated a between-study heterogeneity Q of 0.44 ($p = 0.509$). The between-study heterogeneity when adjusting for whether or not evidence was provided for a functional role of the polymorphism generated a Q of 2.76 ($p = 0.096$). Reporting whether or not supplementary material accompanied the journal article generated a between-study heterogeneity Q of 2.52 ($p = 0.113$).

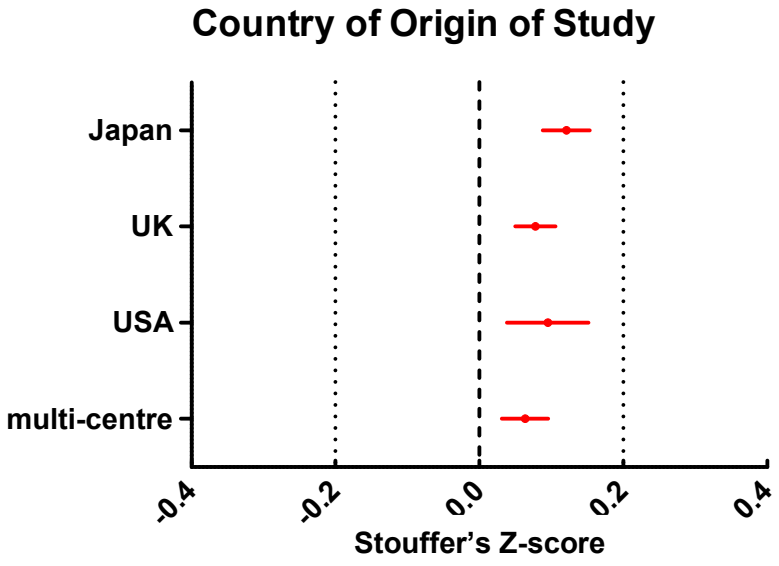
Table 14: Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables): 36 studies

Methodological Characteristic	*Count*	*Stouffer's Z-score*	*95% CI*
Journal			
American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)	2	0.031	(-0.198, 0.257)
BMC Medical Genetics	1	0.210	(0.129, 0.289)
European Journal of Human Genetics	1	0.108	(0.063, 0.152)
Human Genetics	1	0.112	(0.065, 0.158)
Molecular Psychiatry	5	0.042	(-0.061, 0.145)
Nature	5	0.003	(-0.050, 0.057)
Nature Genetics	12	0.104	(0.059, 0.149)
Neuron	1	0.171	(0.120, 0.221)
PLoS Genetics	1	0.071	(0.038, 0.105)
Proceedings of the National Academy of Sciences of the United States of America	1	0.039	(0.025, 0.054)
Science	1	0.139	(0.058, 0.218)
The American Journal of Human Genetics	1	0.110	(0.055, 0.165)
The Lancet Neurology	3	0.041	(-0.189, 0.267)
The New England Journal of Medicine	1	0.047	(0.032, 0.063)
Country of origin of study			
Japan	1	0.121	(0.088, 0.154)
United Kingdom	1	0.078	(0.050, 0.106)
United States of America	8	0.055	(0.042, 0.068)
Multi-centre	26	0.059	(0.053, 0.064)
What disease was the study investigating?			
Alzheimer's disease	4	0.143	(0.035, 0.248)
Amyotrophic lateral sclerosis	6	0.120	(0.073, 0.167)
Attention deficit/hyperactivity disorder	1	-0.086	(-0.123, -0.048)
Autism	2	-0.014	(-0.145, 0.118)
Bipolar disorder	4	0.070	(0.042, 0.097)
Essential tremor	1	0.042	(0.026, 0.058)
Major depressive disorder	1	-0.084	(-0.117, -0.051)
Narcolepsy	2	0.165	(0.126, 0.203)
Parkinson's disease	5	0.059	(-0.014, 0.131)
Progressive supranuclear palsy	1	0.110	(0.055, 0.165)
Restless legs syndrome	2	0.053	(0.032, 0.074)
Schizophrenia	7	0.048	(-0.013, 0.108)
Was the control group:			
Population-based with identified sampling frame	7	0.020	(-0.061, 0.101)
Hospital or other non-population-based specific group (i.e. blood donors)	22	0.077	(0.050, 0.103)
		0.092	(0.009, 0.173)
Mix of population-based and specific group	5	0.127	(0.046, 0.206)
N/A	2		
Were case/control groups related?			
Yes (cryptic relationship)	13	0.068	(0.025, 0.112)
Yes (direct relationship)	2	0.031	(-0.198, 0.257)
No	21	0.074	(0.039, 0.108)
Were cases/controls drawn from the same population?			
Yes	12	0.088	(0.024, 0.151)
No	24	0.059	(0.034, 0.085)

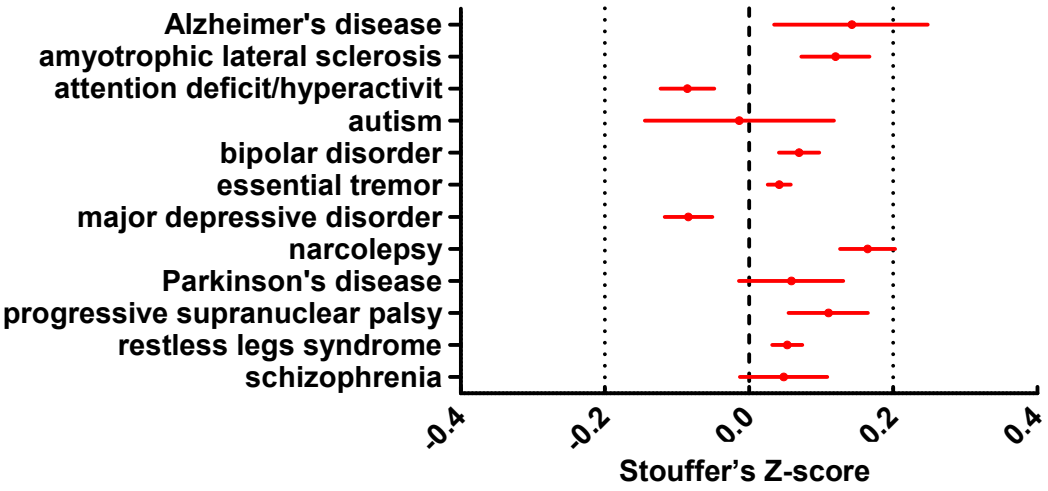
Was the data source:			
Primary collection and analysis	13	0.012	(-0.000, 0.025)
Building on pre-existing data through further collection	21	0.070	(0.064, 0.075)
N/A	2	0.019	(-0.010, 0.047)
Did the study report power/sample size calculations?			
Yes	23	0.044	(0.001, 0.086)
No	13	0.101	(0.077, 0.124)
Was deviation from HWE tested?			
Yes	30	0.070	0.041 (0.098)
No	6	0.066	-0.054 (0.185)
If so, what was the result?			
Non-equilibrium (controls included in further analyses)	3	-0.065	(-0.101, -0.029)
Non-equilibrium (controls excluded from further analyses)	27	0.063	(0.058, 0.068)
Was population stratification assessed?			
Yes	28	0.072	(0.039, 0.105)
No	8	0.073	(0.034, 0.111)
Did the study detect/duplicate previously known results?			
Yes	1	0.087	(0.052, 0.122)
No	35	0.072	(0.045, 0.099)
Was evidence provided for a functional role of the polymorphism?			
Yes	14	0.100	(0.055, 0.144)
No	22	0.054	(0.024, 0.084)
Did supplementary material accompany the journal article?			
Yes	28	0.058	(0.030, 0.087)
No	8	0.130	(0.046, 0.212)

Figure 5: Forest Plots of Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables)

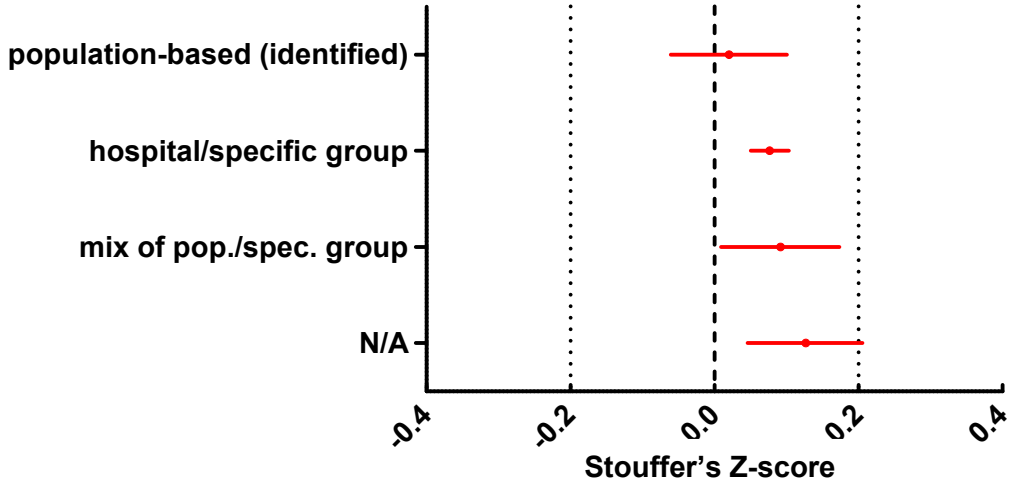




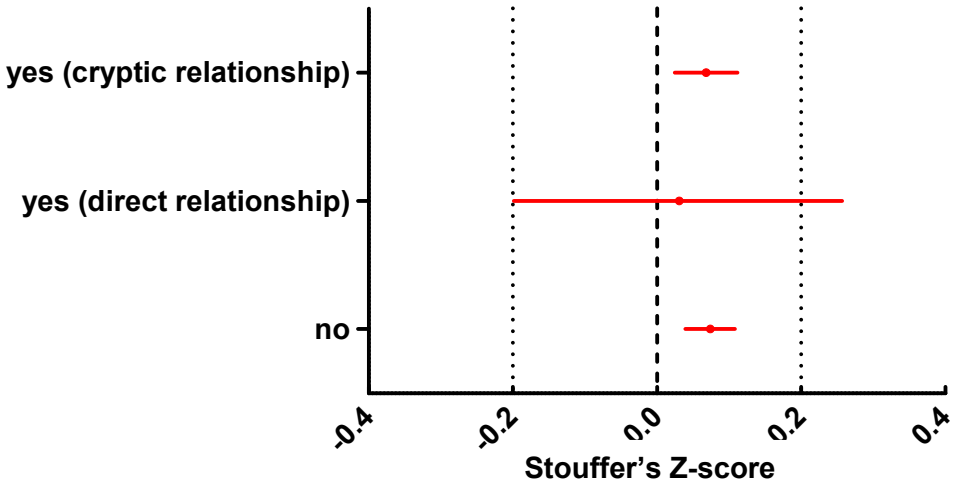
What disease was the study investigating?



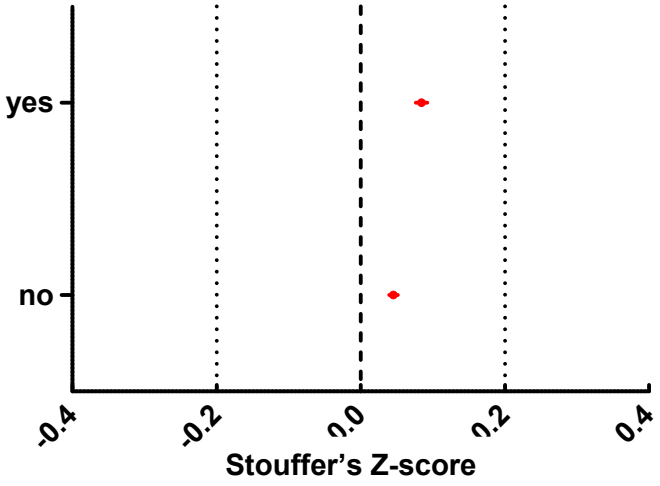
Was the control group:



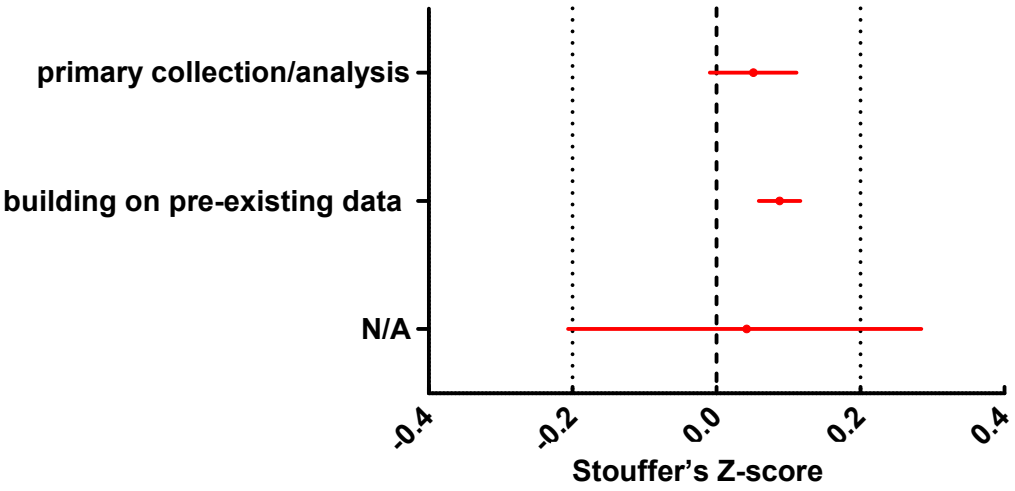
Were case/control groups related?



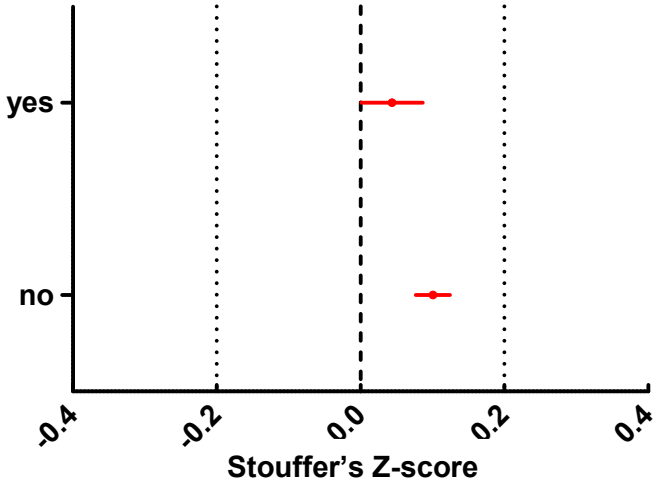
Were cases/controls drawn from the same population?

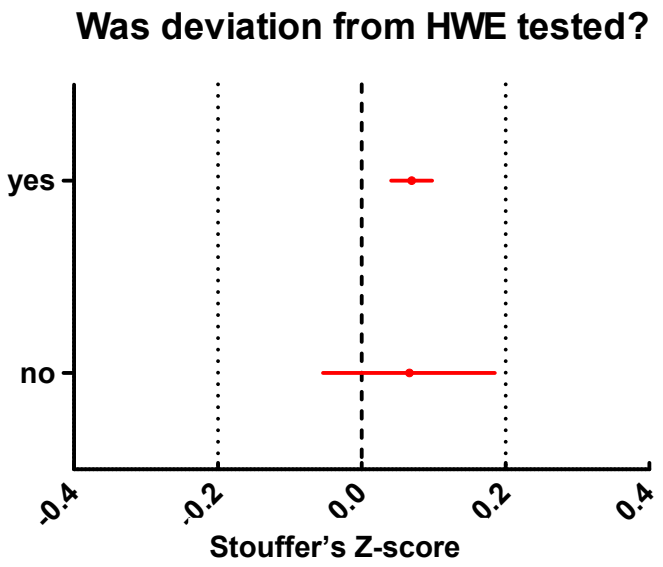


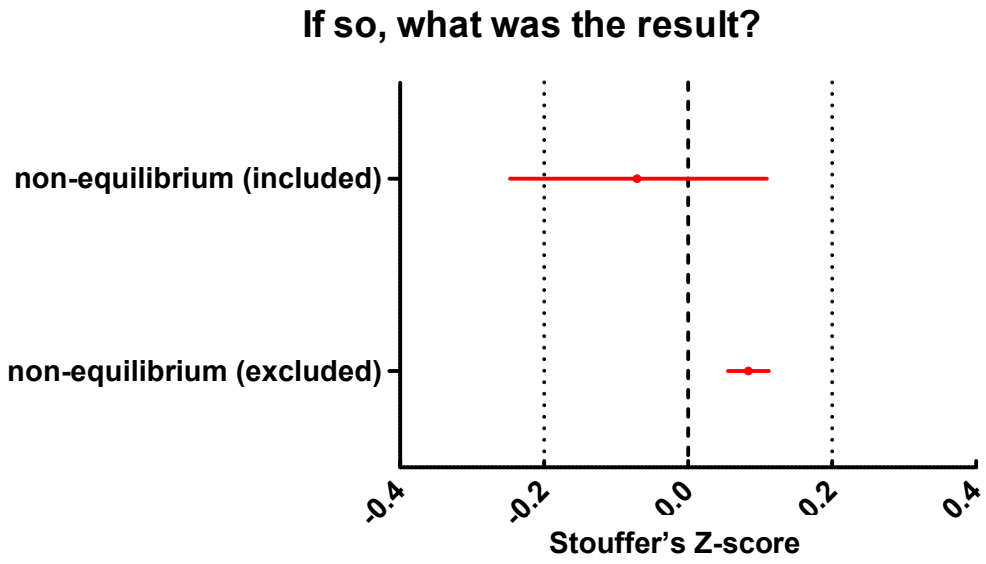
Was the data source:



Did the study report power/sample size calculations?

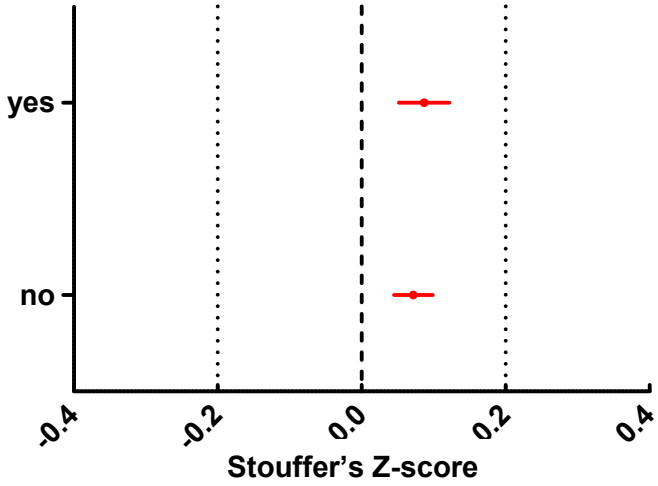



Was deviation from HWE tested?
yes
no
-0.4
-0.2
0.0
0.2
0.4
Stouffer's Z-score

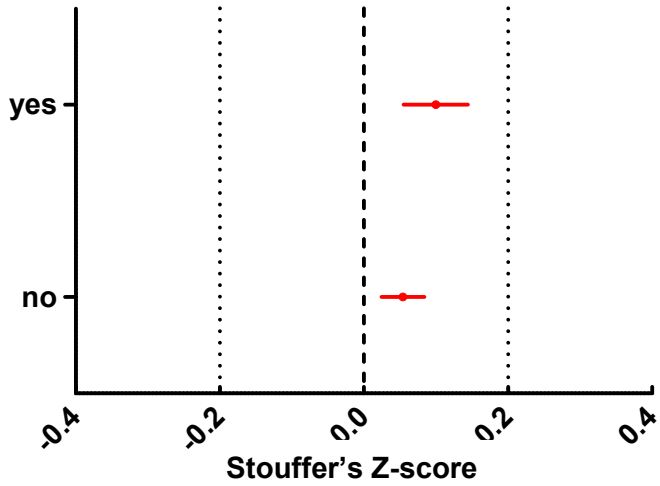

If so, what was the result?
non-equilibrium (included)
non-equilibrium (excluded)
-0.4
-0.2
0.0
0.2
0.4
Stouffer's Z-score

Was population stratification assessed?

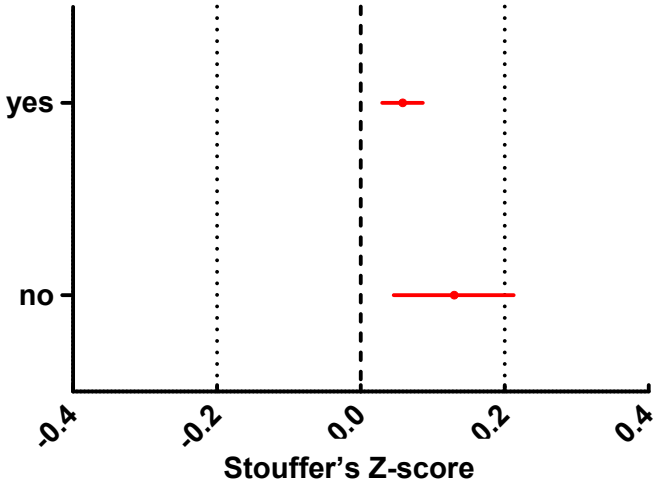
Did the study detect/duplicate previously-known results?



Was evidence provided for a functional role of the polymorpshim?



Did supplementary material accompany the journal article?



Univariate meta-regression under a random effects model using unrestricted maximum likelihood, adjusting for journal impact factor, generated a between-study heterogeneity Q of 0.04 (p = 0.845) (Table 15, Figure 7).

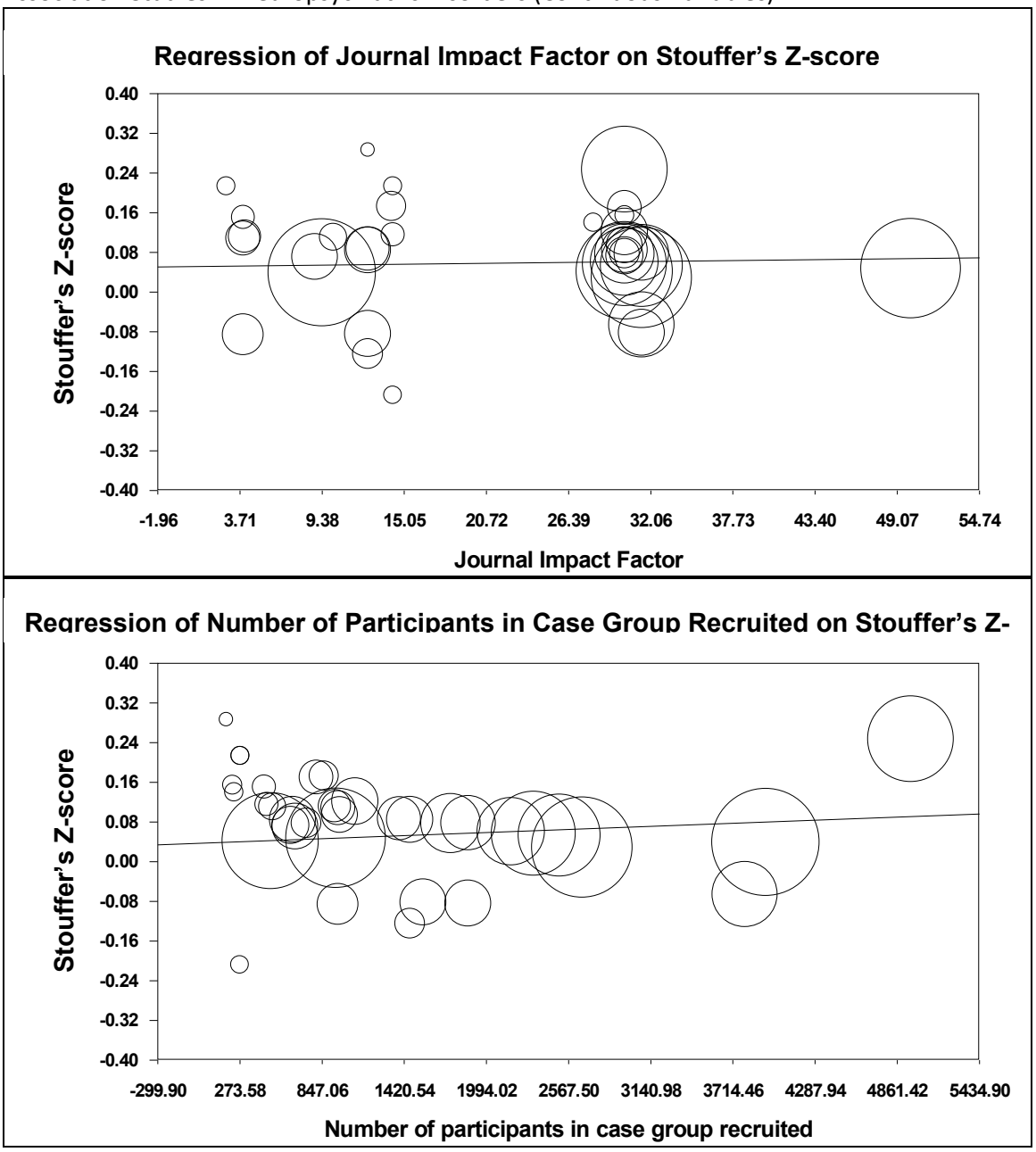
Adjusting for the number of participants recruited for the case group, the between-study heterogeneity Q was 0.76 (p = 0.383). Reporting the number of participants in the case group analyzed generated a between-study heterogeneity Q of 0.72 (p = 0.395). Adjusting for the number of participants recruited for the control group, the between-study heterogeneity Q was 0.60 (p = 0.437). Reporting the number of participants in the control group analyzed generated a between-study heterogeneity Q of 0.54 (p = 0.461).

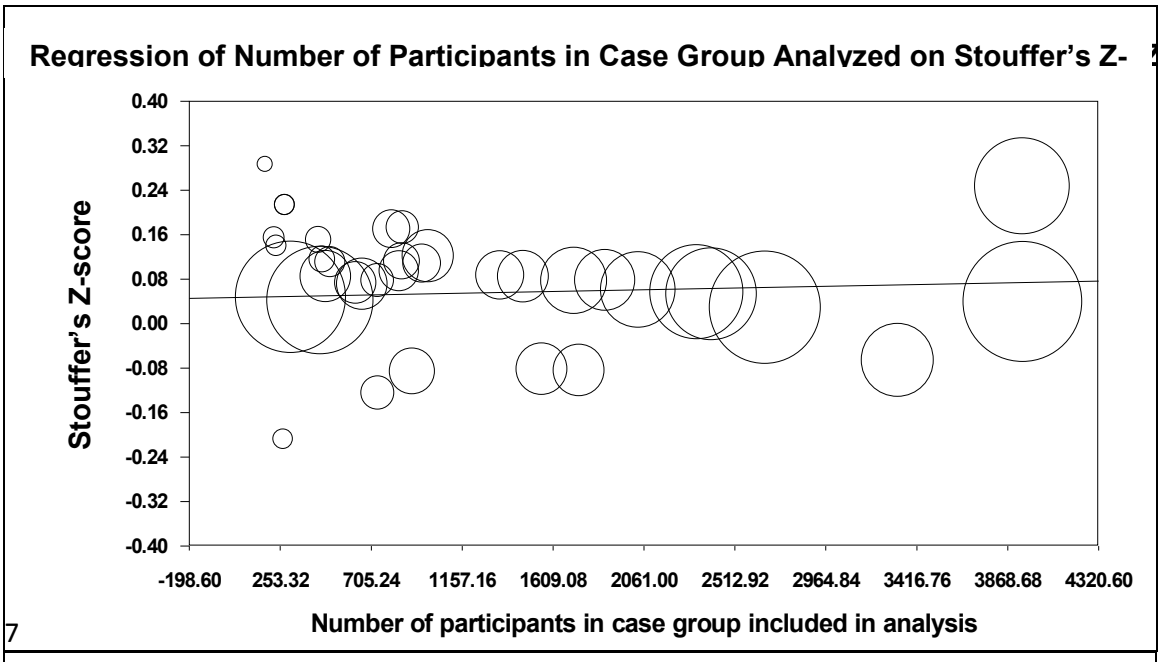
Adjusting for the number of SNPs originally tested, the between-study heterogeneity Q was 0.07 (p = 0.799). Reporting the number of SNPs included in the analysis generated a between-study heterogeneity Q of <0.01 (p = 0.948). Adjusting for the length of reference material included (given that reference material was included), the between-study heterogeneity Q of 5.04 (p = 0.025) accounted for 0.51% of total heterogeneity. Reported Stouffer's Z-scores tended to decrease as the length of reference material included increased (given that reference material was included).

Table 15: Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables)

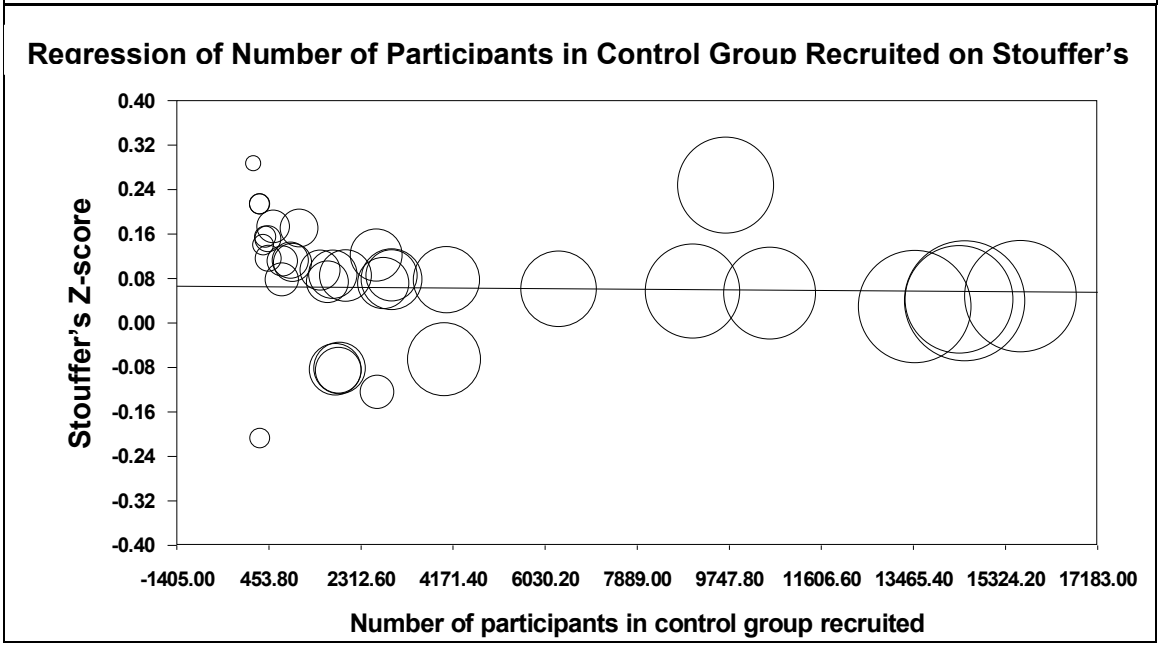
Methodological Characteristic	*Slope*	*95% CI*	*Q-value (p-value)*
Journal Impact Factor	0.00033	(-0.00007, 0.00072)	0.04 (0.845)
Number of study participants in case group recruited	0.00001	(0.00001, 0.00001)	0.76 (0.383)
Number of study participants in case group analyzed	0.00001	(<0.00001, 0.00001)	0.73 (0.395)
Number of study participants in control group recruited	<-0.00001	(<-0.00001, <0.00001)	0.60 (0.437)
Number of study participants in control group analyzed	<-0.00001	(<-0.00001, <-0.00001)	0.54 (0.461)
How many SNPs were originally tested?	<-0.00001	(<-0.00001, <-0.00001)	0.07 (0.799)
How many SNPs were included in the analysis?	<-0.00001	(<-0.00001, <0.00001)	>0.01 (0.948)
If reference material was included, how long was it?	-0.00040	(-0.00048, -0.00033)	5.04 (0.025)

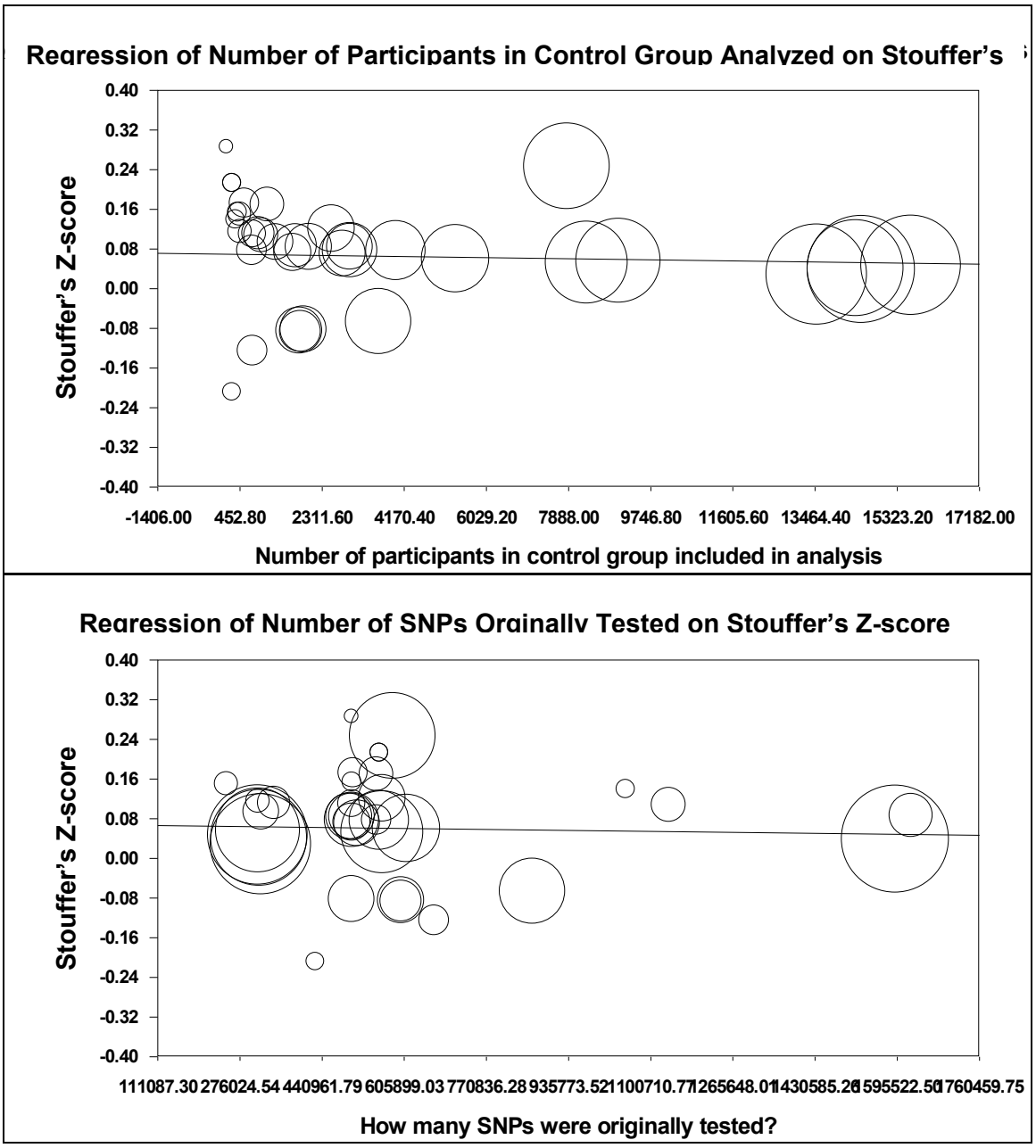
Figure 6: Scatter Plots of Results of Meta-Regression Analysis of Index Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables)

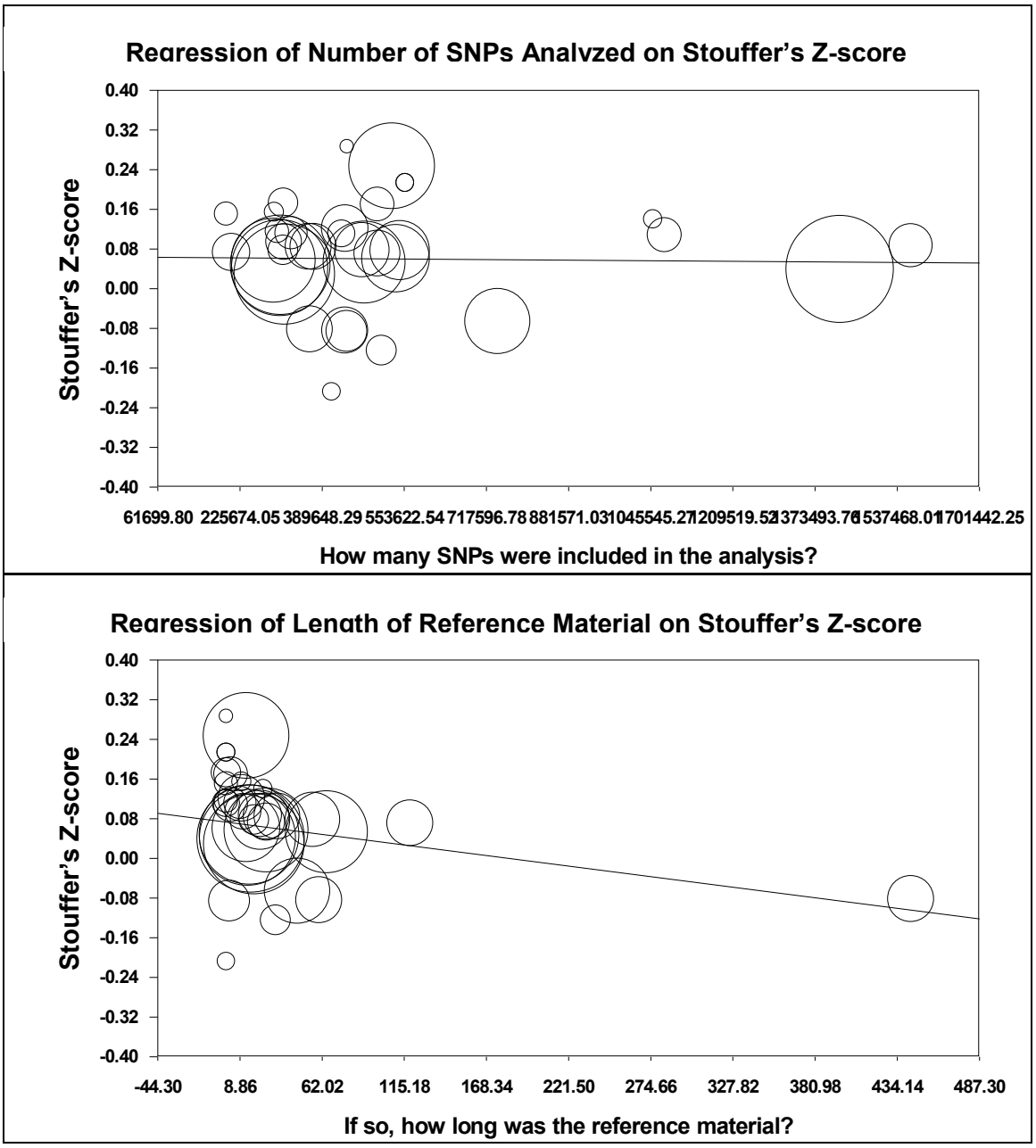



Regression of Number of Participants in Case Group Analyzed on Stouffer's Z-
Stouffer's Z-score
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
-0.16
-0.24
-0.32
-0.40
-198.60
253.32
705.24
1157.16
1609.08
2061.00
2512.92
2964.84
3416.76
3868.68
4320.60
Number of participants in case group included in analysis

7


Regression of Number of Participants in Control Group Recruited on Stouffer's
Stouffer's Z-score
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
-0.16
-0.24
-0.32
-0.40
-1405.00
453.80
2312.60
4171.40
6030.20
7889.00
9747.80
11606.60
13465.40
15324.20
17183.00
Number of participants in control group recruited


Regression of Number of Participants in Control Group Analyzed on Stouffer's
Stouffer's Z-score
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
-0.16
-0.24
-0.32
-0.40
-1406.00
452.80
2311.60
4170.40
6029.20
7888.00
9746.80
11605.60
13464.40
15323.20
17182.00
Number of participants in control group included in analysis
Regression of Number of SNPs Originally Tested on Stouffer's Z-score
Stouffer's Z-score
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
-0.16
-0.24
-0.32
-0.40
111087.30
276024.54
440961.79
605899.03
770836.28
935773.52
1100710.77
1265648.01
1430585.26
1595522.50
1760459.75
How many SNPs were originally tested?


Regression of Number of SNPs Analyzed on Stouffer's Z-score
Stouffer's Z-score
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
-0.16
-0.24
-0.32
-0.40
61699.80
225674.05
389648.29
553622.54
717596.78
881571.03
1045545.27
1209519.52
1373493.76
1537468.01
1701442.25
How many SNPs were included in the analysis?
Regression of Length of Reference Material on Stouffer's Z-score
Stouffer's Z-score
0.40
0.32
0.24
0.16
0.08
0.00
-0.08
-0.16
-0.24
-0.32
-0.40
-44.30
8.86
62.02
115.18
168.34
221.50
274.66
327.82
380.98
434.14
487.30
If so, how long was the reference material?

2.3 Review and Analysis of Reporting of Replication Components of Cancer

Since the objective of the replication components embedded within a genome-wide association study is to identify if the association identified in the index components can be replicated (followed by an estimation of the magnitude of association given one exists), the secondary outcome measure is the reported odds ratio of the replication components. If the replication components did not report an odds ratio within the abstract, then the entire study was reviewed to identify the most extreme reported odds ratio. As certain features of the index and replication components of genome-wide association studies are correlated, some of the features of each journal article will be reviewed in both parts of the analysis.

Descriptive Statistics

Of the fifty-three genome-wide association studies in cancer originally chosen for analysis,[66-118] 45 included at least one replication component (84.9%). The vast majority of the studies investigated replication of effect in 5 or less unique data sets (84.4%, n=38), while the median number of individual replications was 2 (range: 1-12). Of the 45 replication components, only one study failed to replicate their index component results in an independent sample (2.2%), and all of the replication components utilized similar approaches to ascertain case and control groups in both the index and replication components.

The replication components of cancer typically recruited less than 5,000 study participants (42.2%, n=19; Table 16); the median number of study participants recruited across all replications was 5,520 (range: 350-44,331). In 66.7% (n=30) of studies, the control group was recruited from a mix of population-based and specific non-population-based group samples (e.g., blood donors), and in just over half of the studies (55.6%, n=25) replication was performed built on pre-existing data through further

collection. Only two studies did not report the sample handling/genotyping process or methods of quality assurance (4.4%). However, only a minority of studies, 4 of 45 (8.9%), reported on the outcome of quality assurance. Only six studies (13.3%) reported a power/sample size calculation relating to their study size. In almost all of the studies (93.3%, n=42), case and control groups were explicitly reported to be unrelated to each other, and in about two-thirds (62.2%, n=28) case and control groups were reported to have been chosen from the same population. All of the studies used similar DNA extraction methods when collecting data from the case and control groups, however any specifics regarding quantity/quality of DNA were poorly reported overall (as identified through the pilot study process, it was difficult to identify any reasonable way to extract information regarding the specifics of the quantity/quality of DNA used in the replication components). Finally, the median number of SNPs tested was 10 (range: 1 to 121,232); 17 studies (37.8%) sought to replicate ≤10 SNPs, 21 (46.7%) 10-99 SNPs, and 5 (11.1%) ≥100.

Table 16: Design and Conduct of Replication Components of Genome-Wide Association Studies in Cancer

Methodological Characteristic	*Count*	*Percent*
Number of study participants recruited:		
≤5000	19	42.2
5001-9999	16	35.6
≥10,000	10	22.2
Was the control group:		
Population-based with identified sampling frame,	7	15.5
Population-based with unidentified sampling frame,	0	0.00
Hospital or other non-population-based specific group (i.e. blood donors),	8	17.8
Mix of population-based and specific group,	30	66.7
Other/can't tell?	0	0.00
Was the data source:		
Primary collection and analysis,	10	22.2
Building on pre-existing data through further collection,	25	55.6
Secondary analysis of pre-existing data?	0	0.00
N/A	10	22.2
Was the sample handling/genotyping process described?		
Yes	43	95.6
No	2	4.4
Was quality assurance reported?		
Yes	41	91.1
No	4	8.9
Did the study report power/sample size calculations?		
Yes	6	13.3
No	39	86.7
Were case/control groups related?		
Yes (cryptic relationship)	2	4.4
Yes (direct relationship)	1	2.2
No	42	93.3
Were cases/controls drawn from the same population?		
Yes	28	62.2
No	17	37.8
Were DNA extraction methods different for cases/controls?		
Yes	0	0.0
No	45	100.0
How many SNPs were included in the replication component?		
≤10	17	37.8
10-99	21	46.7
≥100	5	11.1
N/A	2	4.4

Hardy-Weinberg equilibrium was assessed in 26.7% of replication components in cancer (n=12). Of these, one-quarter found the SNPs of the control group satisfied equilibrium (n=3), 16.7% included the SNPs of the individuals in the control group that did not satisfy equilibrium (n=2), 16.7% excluded the SNPs of the individuals in the control group that did not satisfy equilibrium (n=2), and 41.7% did not report the result (n=5). The replication components confirmed the results of the original study 91.1% of the time (n=41).

Meta-Analysis

Out of the fifty-three genome-wide association studies in cancer that were chosen for analysis, forty-five performed replication components. Of these forty-five studies, forty-one were included in the meta-analysis. The list of excluded studies along with each study's reason for exclusion is presented in Table 17.

Table 17: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Replication Components of Genome-wide Association Studies in Cancer

Study	*Journal*
Studies Excluded Because of Asymmetric Confidence Limits*	
Capasso 2009[72]	NAT GENET
Eeles 2008[77]	NAT GENET
Shete 2009[98]	NAT GENET
Studies Excluded Because Confidence Limits Not Reported	
Spain 2009[101]	CANCER RES

*confidence limits were considered asymmetric if the reported upper and lower limits were not equivalent to one another

The overall pooled regression coefficient produced from the meta-regression under a random effects model, odds ratio and 95% confidence limit (a summary statistic used to identify effect size), for the replication components of genome-wide association studies in cancer was 1.34 (95% CI 1.25, 1.43). There was considerable heterogeneity between studies (Q 595.84 ($p < 0.001$); I^2 statistic 93.3).

The journal accounted for 9.44% of the total heterogeneity between studies, with a Q of 56.27 ($p = 6.98 \times 10^{-9}$). However, the only journal in which more than two studies were included was *Nature Genetics* (n=31), for which the summary odds ratio was 1.30 (95% CI 1.19, 1.42) (Table 18 and Figure 9). Country of origin of study generated a Q of 5.37 ($p = 0.147$).

The type of disease accounted for 39.20% of total heterogeneity (Q of 233.56 ($p = <1.00 \times 10^{-15}$)). Only the studies that were performed on breast cancer (n=6), colorectal cancer (n=6) and lung cancer (n=5) showed significant heterogeneity. The type of disease that demonstrated the highest odds ratio was testicular germ cell cancer (n=1) (3.08 (95% CI 2.29, 4.14), while follicular lymphoma (n=1) demonstrated the lowest odds ratio (0.43 (95% CI 0.22, 0.84). Reporting on how many replication samples were used generated a between-study heterogeneity Q of 22.68 ($p = 6.96 \times 10^{-3}$), 3.81% of the total heterogeneity.

Adjusting for selection of the control group participants generated a between-study heterogeneity Q of 1.34 ($p = 0.511$). Adjusting for whether or not cases and controls were related, the between-study heterogeneity Q was 0.45 ($p = 0.799$). Reporting on whether or not cases and controls were drawn from the same population generated a between-study heterogeneity Q of 0.47 ($p = 0.792$).

The between-study heterogeneity when adjusting for type of data source generated a Q of 0.69 ($p = 0.709$). Reporting a method to assess quality generated a between-study heterogeneity Q of <0.01 ($p = 0.928$). Univariate meta-regression under a random effects model using unrestricted maximum

likelihood, adjusting for reporting a power/sample size calculation, generated a between-study heterogeneity Q of 3.10 ($p = 0.079$).

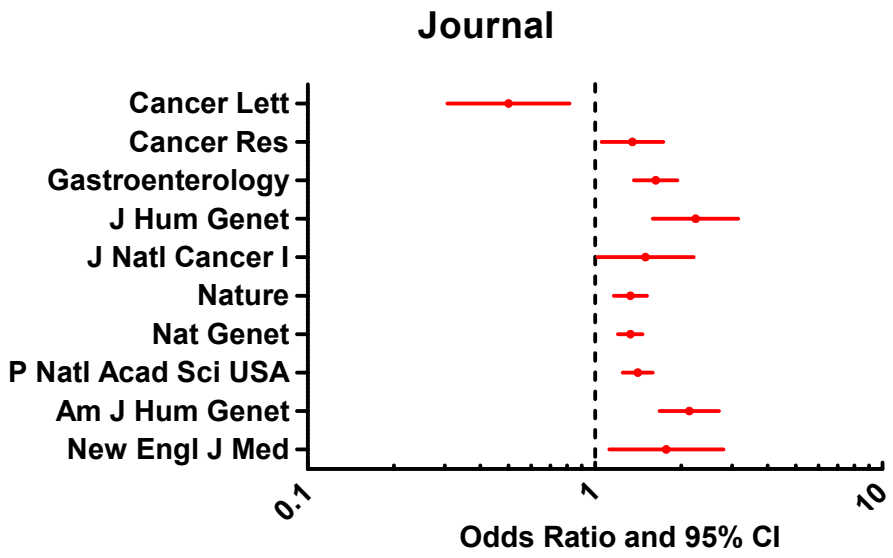
Adjusting for reporting testing for deviation from Hardy-Weinberg equilibrium generated a between-study heterogeneity Q of 6.46 ($p = 0.011$), and showed that between-study heterogeneity accounted for 1.08% of total heterogeneity. Studies that reported testing for deviation from Hardy-Weinberg equilibrium were not significant when compared to those that reported not testing for deviation from Hardy-Weinberg equilibrium. Among the studies that did report on Hardy-Weinberg equilibrium, the between-study heterogeneity generated a Q of 0.35 ($p = 0.950$).

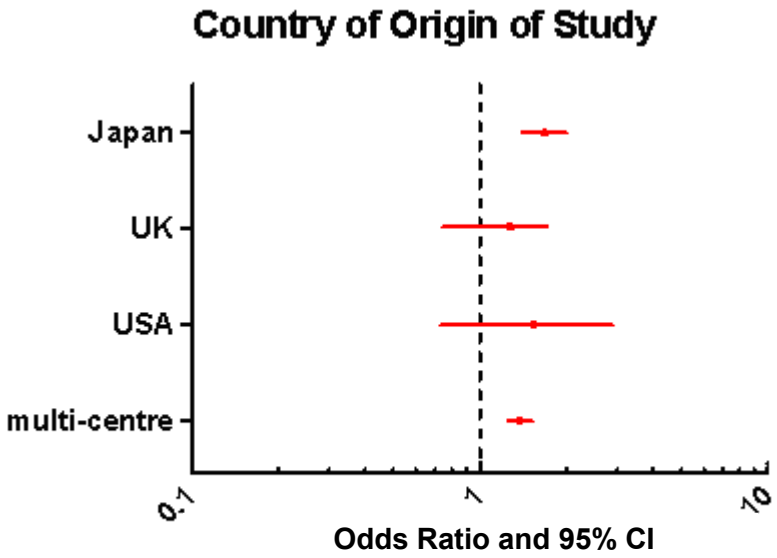
Table 18: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Discrete Variables): 41 studies

Methodological Characteristic	*Count*	*Odds Ratio*	*95% CI*
Journal			
Cancer Letters	1	0.50	(0.31, 0.82)
Cancer Research	1	1.35	(1.05, 1.73)
Gastroenterology	1	1.63	(1.37, 1.94)
Journal of Human Genetics	1	2.24	(1.59, 3.15)
Journal of the National Cancer Institute	2	1.50	(1.02, 2.21)
Nature	31	1.26	(1.23, 1.29)
Nature Genetics	1	1.30	(1.19, 1.42)
Proceedings of the National Academy of Sciences of the United States of America	1	1.41	(1.25, 1.59)
The American Journal of Human Genetics	1	2.13	(1.68, 2.70)
The New England Journal of Medicine	1	1.77	(1.12, 2.80)
Country of Origin of Study			
Japan	1	1.63	(1.37, 1.94)
United Kingdom	3	1.32	(1.24, 1.42)
United States of America	4	1.17	(0.79, 1.72)
Multi-centre	33	1.49	(0.77, 2.88)
What disease was the study investigating?			
Breast cancer	6	1.36	(1.17, 1.58)
Chronic lymphocytic leukemia	1	1.65	(1.40, 1.95)
Colorectal cancer	6	1.08	(0.92, 1.25)
Cutaneous basal cell carcinoma	1	1.33	(1.11, 1.59)
Esophageal cancer	1	1.63	(1.37, 1.94)
Follicular lymphoma	1	0.43	(0.22, 0.84)
Glioma	1	0.48	(0.32, 0.72)
Lung cancer	5	1.24	(1.08, 1.43)
Melanoma	2	1.76	(1.37, 2.26)
Nasopharyngeal carcinoma	2	2.17	(1.78, 2.63)
Neuroblastoma	1	1.77	(1.12, 2.80)
Ovarian Cancer	1	0.88	(0.81, 0.95)
Pancreatic cancer	1	1.15	(1.05, 1.26)
Prostate cancer	8	1.46	(1.34, 1.61)
Testicular germ cell cancer	1	3.08	(2.29, 4.14)
Thyroid cancer	1	1.81	(1.45, 2.26)
Urinary bladder cancer	2	1.42	(1.19, 1.70)
How many replications were performed?			
1	8	1.31	(1.11, 1.55)
2	13	1.46	(1.31, 1.63)
3	6	1.35	(1.03, 1.78)
4	4	1.28	(0.69, 2.35)
5	3	1.41	(1.29, 1.53)
6	1	1.35	(1.05, 1.73)
7	2	1.46	(1.26, 1.69)
8	2	1.02	(0.75, 1.40)
9	1	0.56	(0.36, 0.86)
12	1	1.40	(1.02, 1.93)
Was the control group:			
Population-based with identified sampling frame	6	1.53	(1.16, 2.02)
Hospital or other non-population-based specific group	6	1.32	(1.01, 1.73)

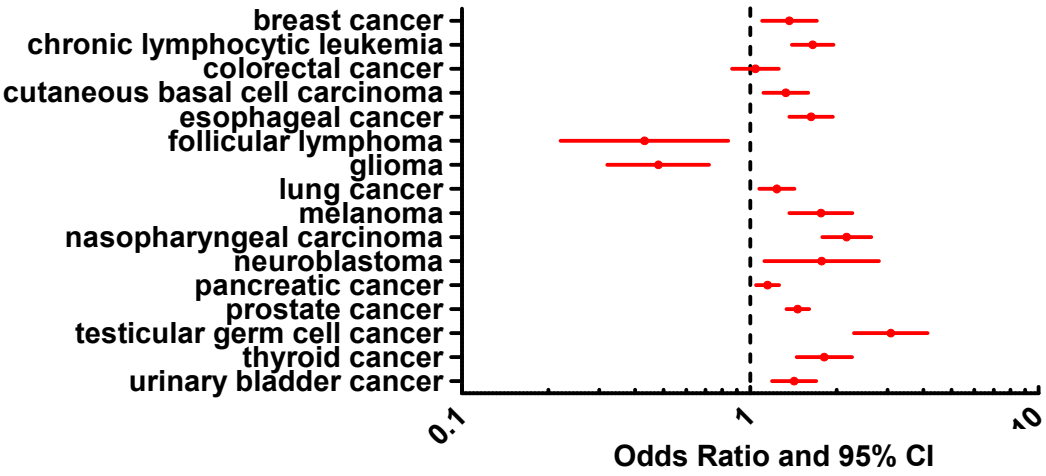
(i.e. blood donors)			
Mix of population-based and specific group	29	1.29	(1.29, 1.39)
Were case/control groups related?			
Yes (cryptic relationship)	2	1.60	(0.93, 2.76)
Yes (direct relationship)	1	1.33	(1.11, 1.59)
No	38	1.32	(1.23, 1.42)
Were cases/controls drawn from the same population?			
Yes	27	1.35	(1.24, 1.46)
No	13	1.27	(1.09, 1.49)
N/A	1	1.40	(1.02, 1.93)
Was the data source:			
Primary collection and analysis	9	1.28	(1.02, 1.60)
Building on pre-existing data through further collection	23	1.33	(1.22, 1.46)
N/A	9	1.39	(1.26, 1.54)
Was quality assurance reported?			
Yes	38	1.34	(1.25, 1.44)
No	3	1.21	(1.09, 1.35)
Did the study report power/sample size calculations?			
Yes	5	1.12	(0.91, 1.40)
No	36	1.38	(1.28, 1.49)
Was deviation from HWE tested?			
Yes	6	1.14	(0.98, 1.31)
No	35	1.42	(1.30, 1.55)
If so, what was the result?			
Equilibrium	2	0.97	(0.28, 3.42)
Non-equilibrium (controls included in further analyses)	2	1.12	(0.29, 4.31)
Non-equilibrium (controls excluded from further analyses)	2	1.37	(1.11, 1.68)

Figure 7: Forest Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Discrete Variables)

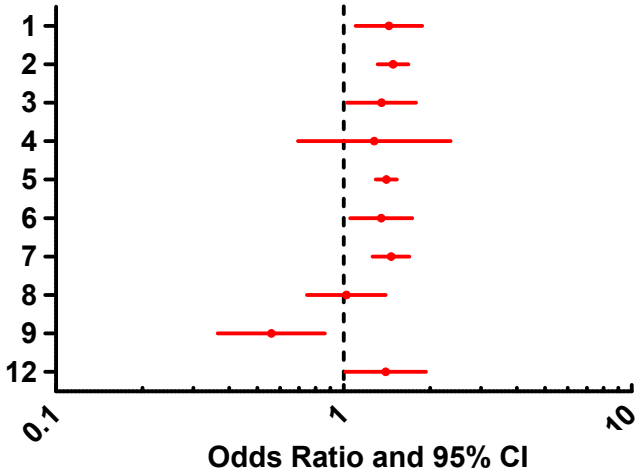




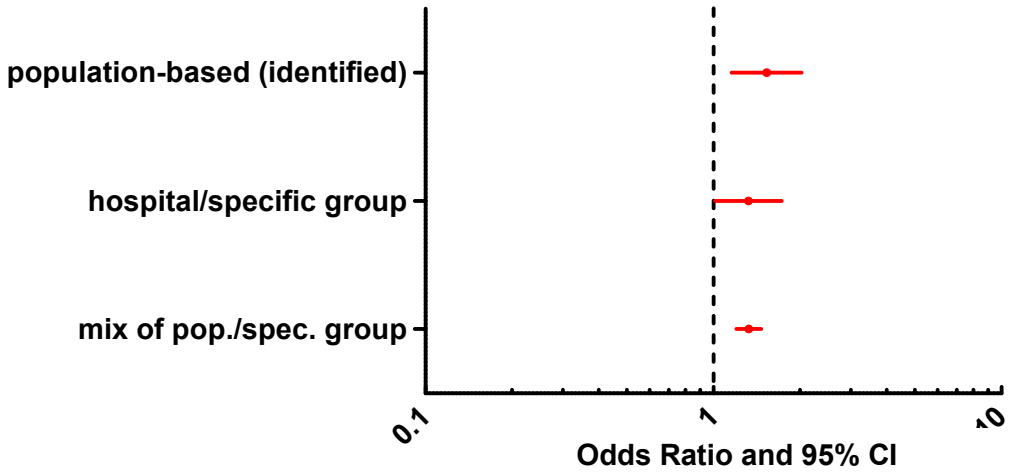
What disease was the study investigating?



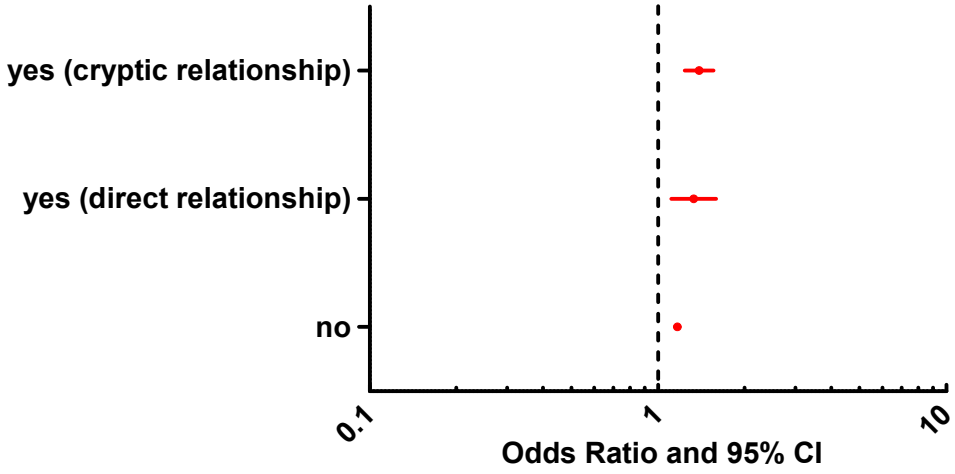
How many replication samples were used?



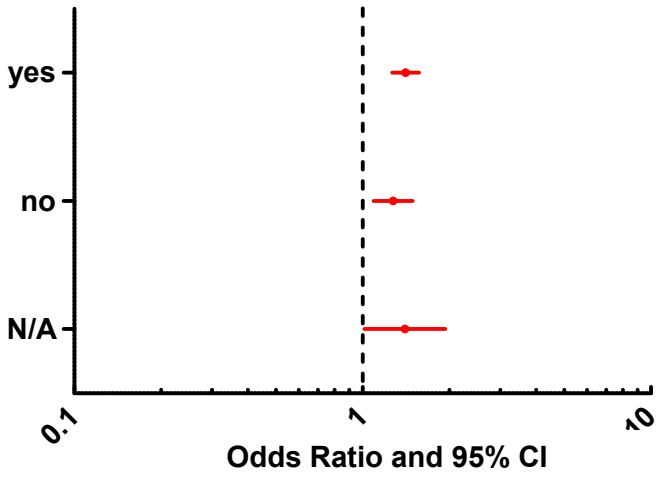
Was the control group?



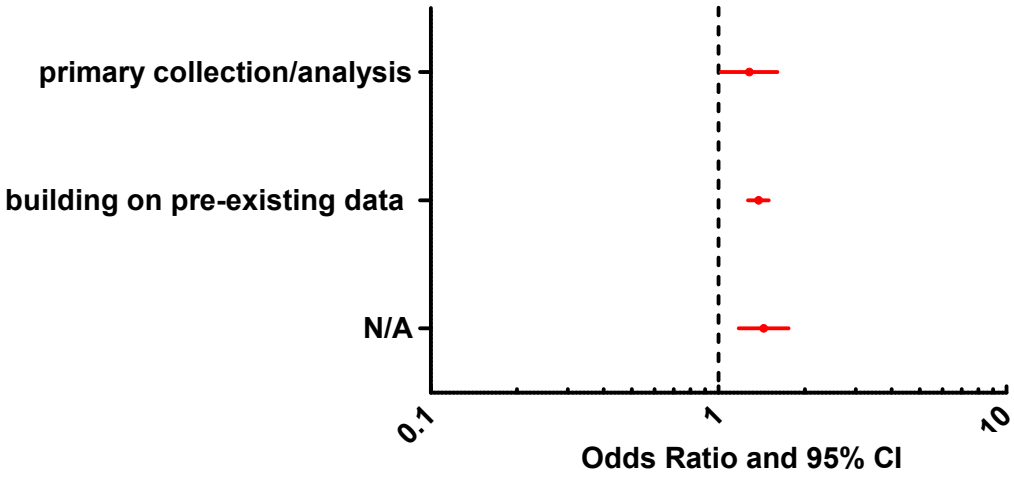
Were case/control groups related?



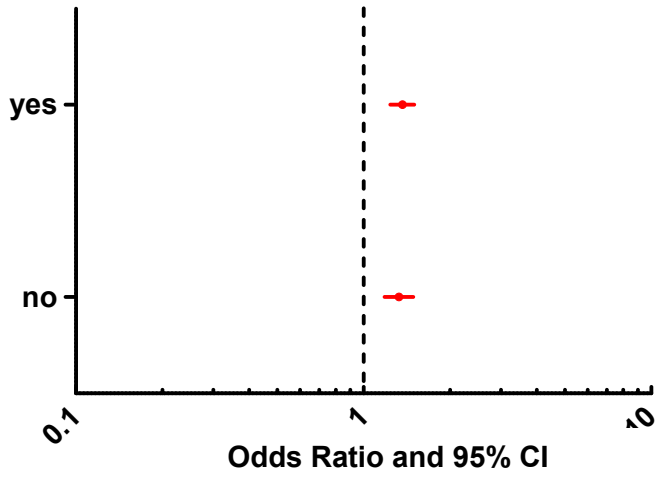
Were cases/controls drawn from the same population?



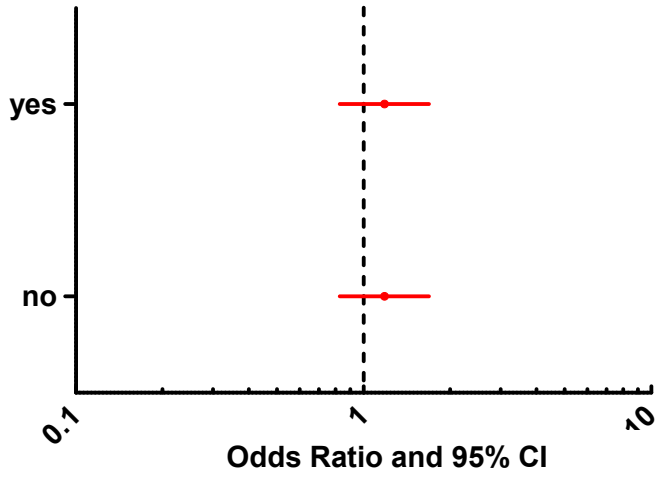
Was the data source:



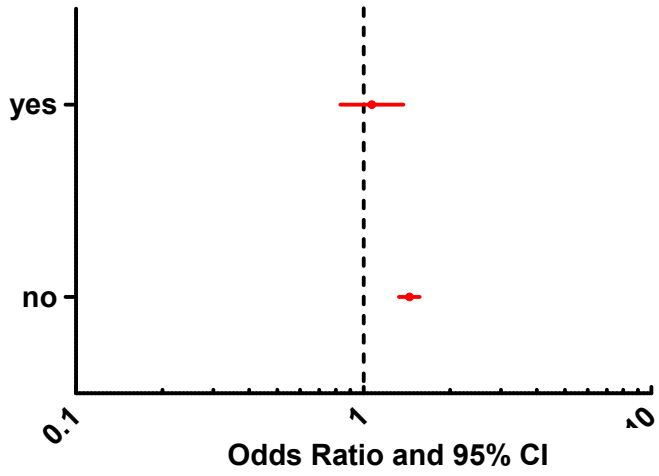
Was quality assurance reported?



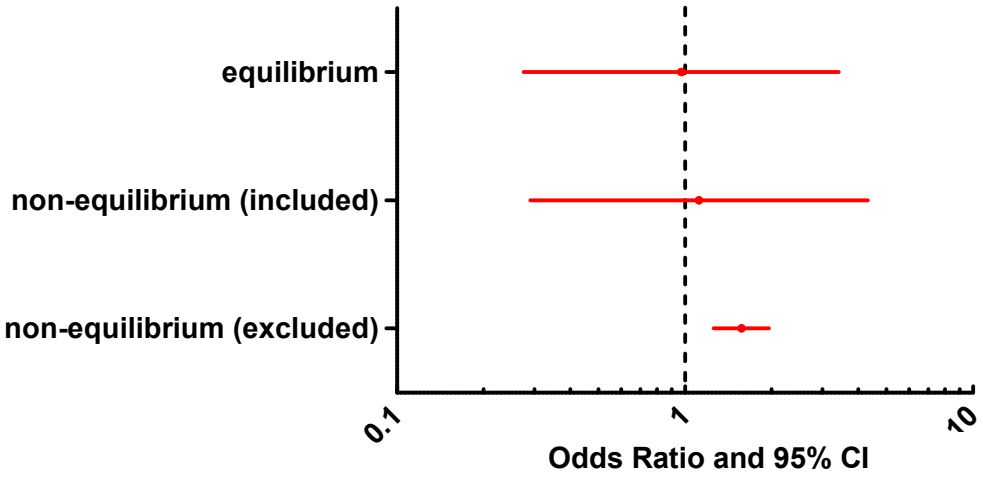
Did the study report power/sample size calculations?



Was deviation from HWE tested?



If so, what was the result?

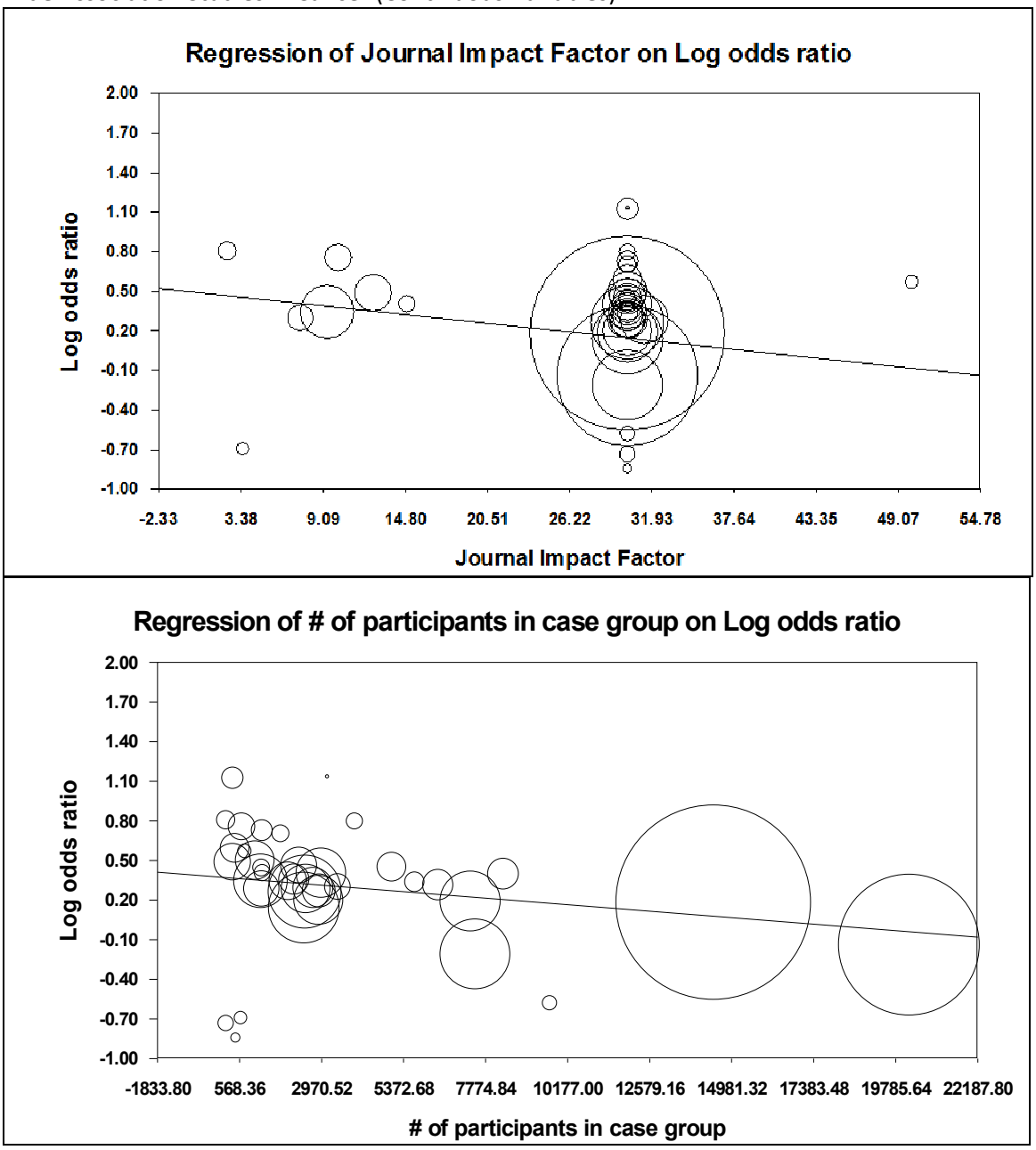


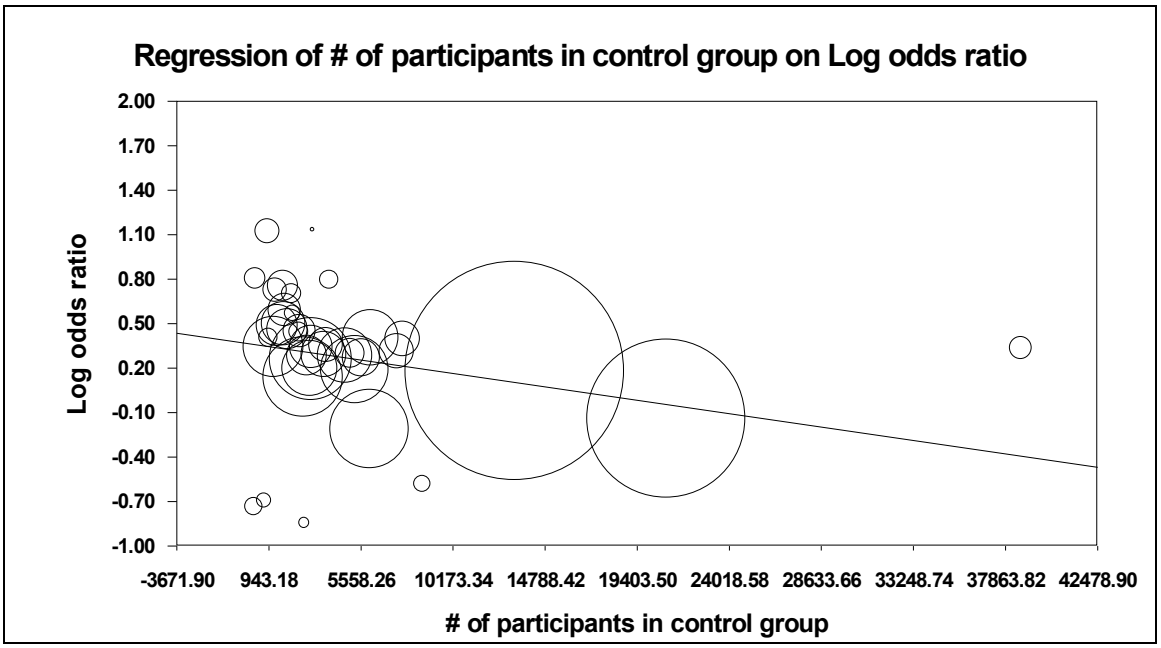
Univariate meta-regression under a random effects model using unrestricted maximum likelihood, adjusting for journal impact factor, generated a between-study heterogeneity Q of 0.03 (p = 0.856) (Table 19, Figure 10). Adjusting for the number of participants recruited for the case group, the between-study heterogeneity Q of 5.61 (p = 0.018) accounted for 1.09% of total heterogeneity. Reported odds ratios tended to decrease as the number of participants in the case group increased. Reporting the number of participants in the control group generated a between-study heterogeneity Q of 1.38 (p = 0.240).

Table 19: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Continuous Variables)

Methodological Characteristic	*Slope*	*95% CI*	*Q-value (p-value)*
Journal Impact Factor	-0.01156	(-0.01546, -0.00765)	0.03 (0.856)
Number of study participants in case group	-0.00002	(-0.00002, -0.00002)	5.61 (0.018)
Number of study participants in control group	-0.00002	(-0.00002, -0.00002)	1.38 (0.240)

Figure 8: Scatter Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Cancer (Continuous Variables)




Regression of # of participants in control group on Log odds ratio
Log odds ratio
2.00
1.70
1.40
1.10
0.80
0.50
0.20
-0.10
-0.40
-0.70
-1.00
-3671.90
943.18
5558.26
10173.34
14788.42
19403.50
24018.58
28633.66
33248.74
37863.82
42478.90
# of participants in control group

2.4 Review and Analysis of Reporting of Replication Components of Neuropsychiatric Disorders

Since the objective of the replication components embedded within a genome-wide association study is to identify if the association identified in the index components can be replicated (followed by an estimation of the magnitude of association given one exists), the secondary outcome measure is the reported odds ratio of the replication components. If the replication components did not report an odds ratio within the abstract, then the entire study was reviewed to identify the most extreme reported odds ratio. As certain features of the index and replication components of the genome-wide association study are correlated, some of the features of each journal article will be reviewed in both parts of the analysis.

Descriptive Statistics

Of the seventy-four genome-wide association studies in neuropsychiatric disorders originally chosen for analysis, 39 included at least one replication component (52.7%). The vast majority of the studies investigated replication of effect in 5 or less unique data sets (97.4%, n=38), while the median number of individual replications was 2 (range: 1-8). Of the 39 replication components, only one study failed to replicate their index component results in an independent sample (2.6%), and all of the replication components utilized similar approaches to ascertain case and control groups in both the index and replication components.

The replication components of neuropsychiatric disorders typically recruited less than 5,000 study participants (71.8%, n=28, Table 20); the median number of study participants recruited across all replications was 2,171 (range: 458-20,554). In 33.3% (n=13) of studies, the control group was recruited from a hospital or other specific non-population-based group sample (e.g., blood donors), and in just over half of the studies (51.3%, n=20) replication was performed built on pre-existing data through

further collection. Only two studies did not report the sample handling/genotyping process or methods of quality assurance (5.1%). However, only a minority of studies, 9 of 39 (23.1%), reported on the outcome of quality assurance. Only 16 studies (41.0%) reported a power/sample size calculation for replication. In almost all of the studies (92.3%, n=36), case and control groups were explicitly reported to be unrelated to each other, and in just under three-quarters (71.8%, n=28) case and control groups were reported to have been not chosen from the same population. The vast majority of studies used similar DNA extraction methods when collecting data from the case and control groups (94.9%, n=2), however any specifics regarding quantity/quality of DNA were poorly reported overall (as identified through the pilot study process, it was difficult to identify any reasonable way to extract information regarding the specifics of the quantity/quality of DNA used in the replication components). Finally, the median number of SNPs tested was 37.5 (range: 1 to 287,522); 7 studies (18.0%) sought to replicate ≤10 SNPs, 16 (41.0%) 10-99 SNPs, and 15 (38.5%) ≥100.

Table 20: Design and Conduct of Replication Components of Genome-Wide Association Studies in Neuropsychiatric Disorders

Methodological Characteristic	*Count*	*Percent*
Number of study participants recruited:		
≤5000	28	71.8
5001-9999	6	15.4
≥10,000	5	12.8
Was the control group:		
Population-based with identified sampling frame,	9	23.1
Population-based with unidentified sampling frame,	0	0.0
Hospital or other non-population-based specific group (i.e. blood donors),	13	33.3
Mix of population-based and specific group,	11	28.2
Other/can't tell?	6	15.4
Was the data source:		
Primary collection and analysis,	9	23.1
Building on pre-existing data through further collection,	20	51.3
Secondary analysis of pre-existing data?	0	0.0
N/A	10	25.6
Was the sample handling/genotyping process described?		
Yes	37	94.9
No	2	5.1
Was quality assurance reported?		
Yes	30	76.9
No	9	23.1
Did the study report power/sample size calculations?		
Yes	16	41.0
No	23	59.0
Were case/control groups related?		
Yes (cryptic relationship)	3	7.7
Yes (direct relationship)	0	0.0
No	36	92.3
Were cases/controls drawn from the same population?		
Yes	11	28.2
No	28	71.8
Were DNA extraction methods different for cases/controls?		
Yes	2	5.1
No	37	94.9
How many SNPs were included in the replication component?		
≤10	7	18.0
10-99	16	41.0
≥100	15	38.5
N/A	1	2.5

Hardy-Weinberg equilibrium was assessed in 25.6% of replication components in neuropsychiatric disorders (n=10). Of these, half found the SNPs of the control group satisfied equilibrium (n=5), 30.0% included the SNPs of the individuals in the control group that did not satisfy equilibrium (n=3), and 20.0% excluded the SNPs of the individuals in the control group that did not satisfy equilibrium (n=2). The replication components confirmed the results of the original study 74.4% of the time (n=29).

Meta-Analysis

Out of the seventy-four genome-wide association studies in neuropsychiatric disorders that were chosen for analysis, thirty-nine performed replication components. Of these thirty-nine studies, twenty were included in the meta-analysis. The list of excluded studies along with each study's reason for exclusion is presented in Table 21.

-

Table 21: List of Excluded Studies Along with Reason for Exclusion in the Meta-Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders

Study	*Journal*
Studies Excluded Because of Asymmetrical Confidence Limits*	
Hallmayer 2009	Nature Genetics
Stefansson 2007	The New England Journal of Medicine
van Es 2009	Nature Genetics
Studies Excluded Because Odds Ratios Missing	
Beecham 2009	The American Journal of Human Genetics
Bertram 2008	The American Journal of Human Genetics
Hattori 2009	American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)
Huentelmann 2007	Human Molecular Genetics
Landers 2009	Proceedings of the National Academy of Sciences of the United States of America
Latourelle 2009	BMC Medical Genetics
Mead 2008	The Lancet Neurology
Sullivan 2009	Molecular Psychiatry
van den Oord 2008	Archives of General Psychiatry
Studies Excluded Because Confidence Limits Missing	
Need 2009	PLoS Genetics
O'Donovan 2008	Nature Genetics
Simon-Sanchez 2009	Nature Genetics
Sklar 2008	Molecular Psychiatry
Smith 2009	Molecular Psychiatry
Wang 2009	Nature
Weiss 2009	Nature

*confidence limits were considered asymmetric if the reported upper and lower limits were not equivalent to one another

The overall pooled regression coefficient produced from the meta-regression under a random effects model, odds ratio and 95% confidence limit (a summary statistic used to identify effect size), for the replication components of genome-wide association studies in neuropsychiatric disorders was 1.43 (95% CI 1.30, 1.57). There was considerable heterogeneity between studies (Q 145.52 ($p < 0.001$); I^2 statistic 86.9).

The journal accounted for 13.9% of the total heterogeneity between studies, with a Q of 20.15 ($p = 0.028$). However, the only journal in which more than two studies were collected from was *Nature Genetics* (n=9), which generated a summary odds ratio of 1.43 (95% CI 1.19, 1.71) (Table 22 and Figure 12). Country of origin of study generated a Q of 0.82 ($p = 0.662$).

The type of disease accounted for 12.9% of total heterogeneity (Q of 18.73 ($p = 0.016$)). Only the studies that were performed on Alzheimer's disease (n=4), amyotrophic lateral sclerosis (n=6) and Parkinson's disease (n=2) showed significant heterogeneity. The type of disease that demonstrated the highest odds ratio was autism (n=1) (2.90 (95% CI 1.03, 8.16), while schizophrenia demonstrated the lowest odds ratio (n=2) (1.29 (95% CI 1.18, 1.41). Reporting on how many replication samples were used generated a between-study heterogeneity Q of 0.61 ($p = 0.895$).

Adjusting for selection of the control group participants generated a between-study heterogeneity Q of 1.97 ($p = 0.579$). Adjusting for whether or not-cases and controls were related, the between-study heterogeneity Q was <0.01 ($p = 0.923$). Reporting on whether or not cases and controls were drawn from the same population generated a between-study heterogeneity Q of 1.18 ($p = 0.278$).

The between-study heterogeneity when adjusting for type of data source generated a Q of 4.52 ($p = 0.211$). Reporting a method to assess quality generated a between-study heterogeneity Q of 0.05 ($p = 0.829$). Univariate meta-regression under a random effects model using unrestricted maximum

likelihood, adjusting for reporting a power/sample size calculation, generated a between-study heterogeneity Q of 2.42 ($p = 0.120$).

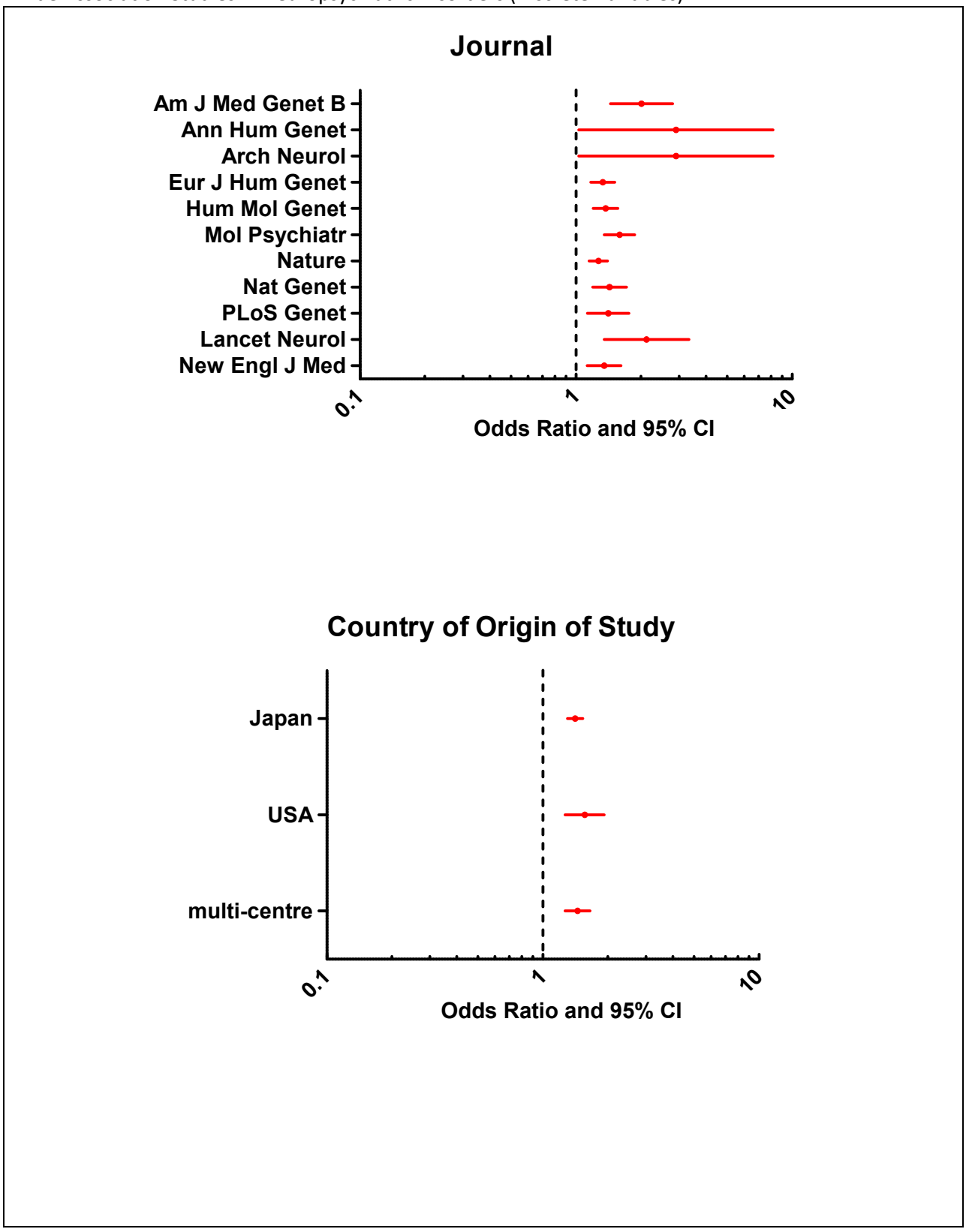
Adjusting for reporting testing for deviation from Hardy-Weinberg equilibrium generated a between-study heterogeneity Q of 3.62 ($p = 0.057$). Among the studies that did report on Hardy-Weinberg equilibrium, the between-study heterogeneity generated a Q of 3.28 ($p = 0.351$).

Table 22: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables): 20 studies

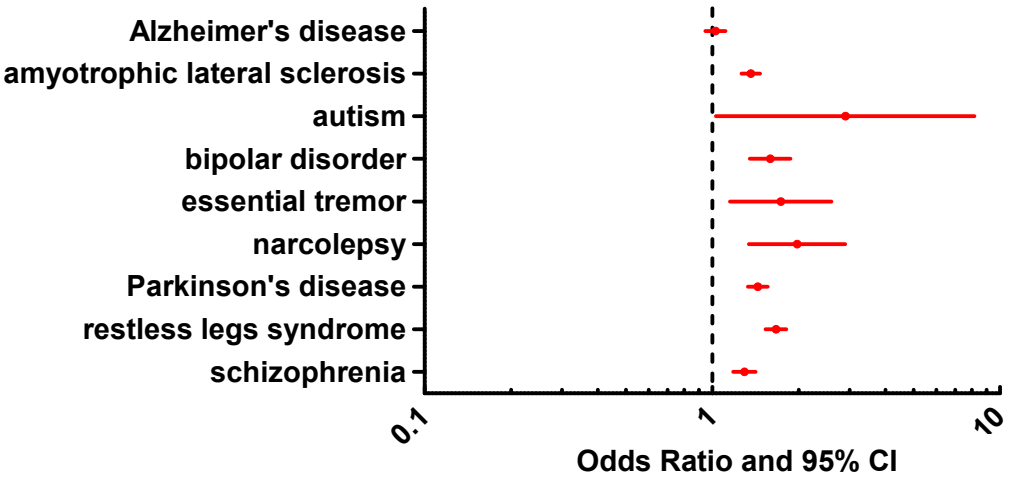
Methodological Characteristic	*Count*	*Odds Ratio*	*95% CI*
Journal			
American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)	1	2.01	(1.44, 2.80)
Annals of Human Genetics	1	2.90	(1.03, 8.16)
Archives of Neurology	1	2.90	(1.03, 8.16)
European Journal of Human Genetics	1	1.33	(1.17, 1.52)
Human Molecular Genetics	2	1.25	(1.07, 1.46)
Molecular Psychiatry	1	1.59	(1.35, 1.87)
Nature	1	1.27	(1.15, 1.40)
Nature Genetics	9	1.43	(1.19, 1.71)
PLoS Genetics	1	1.41	(1.13, 1.76)
The Lancet Neurology	1	2.12	(1.35, 3.33)
The New England Journal of Medicine	1	1.35	(1.13, 1.62)
Country of Origin of Study			
Japan	1	1.41	(1.30, 1.53)
United States of America	4	1.56	(1.27, 1.93)
Multi-centre	15	1.41	(1.26, 1.58)
What disease was the study investigating?			
Alzheimer's disease	4	1.30	(0.92, 1.84)
Amyotrophic lateral sclerosis	6	1.32	(1.19, 1.45)
Autism	1	2.90	(1.03, 8.16)
Bipolar disorder	1	1.59	(1.35, 1.87)
Essential tremor	1	1.73	(1.15, 2.60)
Narcolepsy	1	1.97	(1.34, 2.90)
Parkinson's disease	2	1.62	(1.15, 2.28)
Restless legs syndrome	2	1.64	(1.43, 1.88)
Schizophrenia	2	1.29	(1.18, 1.41)
How many replications were performed?			
1	5	1.65	(1.05, 2.57)
2	5	1.44	(1.30, 1.60)
3	4	1.39	(1.15, 1.67)
4	6	1.40	(1.26, 1.56)
Was the control group:			
Population-based with identified sampling frame	7	1.59	(1.30, 1.94)
Hospital or other non-population-based specific group (i.e. blood donors)	6	1.39	(1.31, 1.48)
Mix of population-based and specific group	6	1.32	(1.07, 1.63)
N/A	1	1.37	(1.20, 1.56)
Were case/control groups related?			
Yes (cryptic relationship)	1	1.45	(1.16, 1.81)
No	19	1.43	(1.30, 1.58)
Were cases/controls drawn from the same population?			
Yes	8	1.33	(1.10, 1.61)
No	12	1.50	(1.35, 1.67)
Was the data source:			
Primary collection and analysis	5	1.49	(1.28, 1.73)
Building on pre-existing data through further collection	10	1.34	(1.18, 1.53)
N/A	5	1.53	(1.28, 1.83)
Was quality assurance reported?			

Yes	18	1.44	(1.33, 1.56)
No	2	1.32	(0.60, 2.91)
Did the study report power/sample size calculations?			
Yes	10	1.33	(1.17, 1.52)
No	10	1.52	(1.37, 1.69)
Was deviation from HWE tested?			
Yes	6	1.26	(1.07, 1.48)
No	14	1.50	(1.38, 1.63)
If so, what was the result?			
Equilibrium	3	1.19	(0.85, 1.67)
Non-equilibrium (controls included in further analyses)	2	1.35	(1.00, 1.82)
Non-equilibrium (controls excluded from further analyses)	1	1.31	(1.08, 1.58)

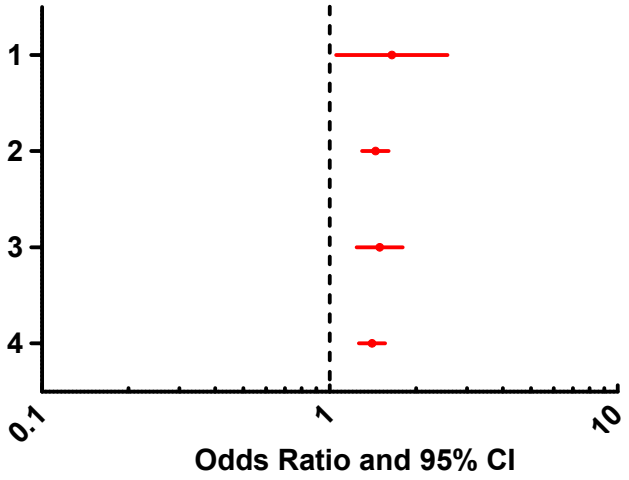
Figure 9: Forest Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Discrete Variables)



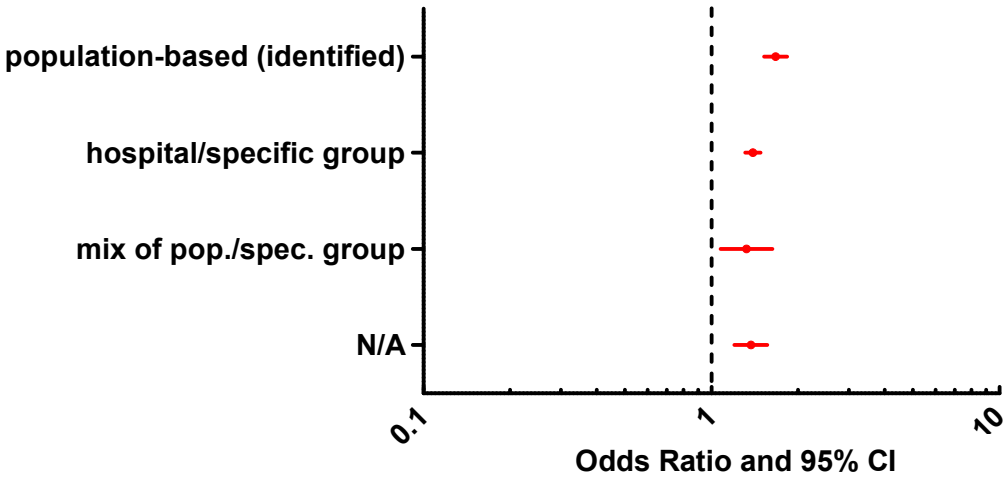
What disease was the study investigating?



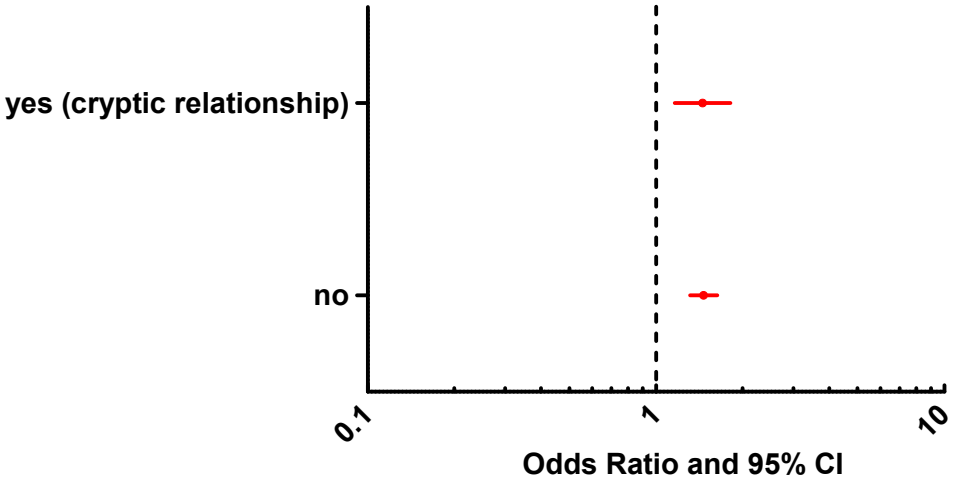
How many replication samples were used?



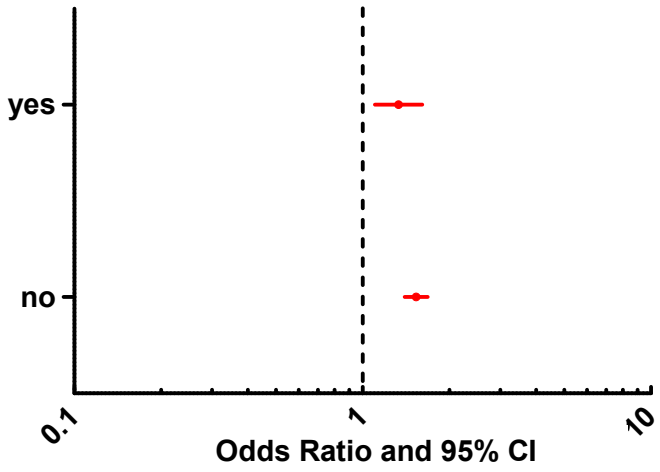
Was the control group:



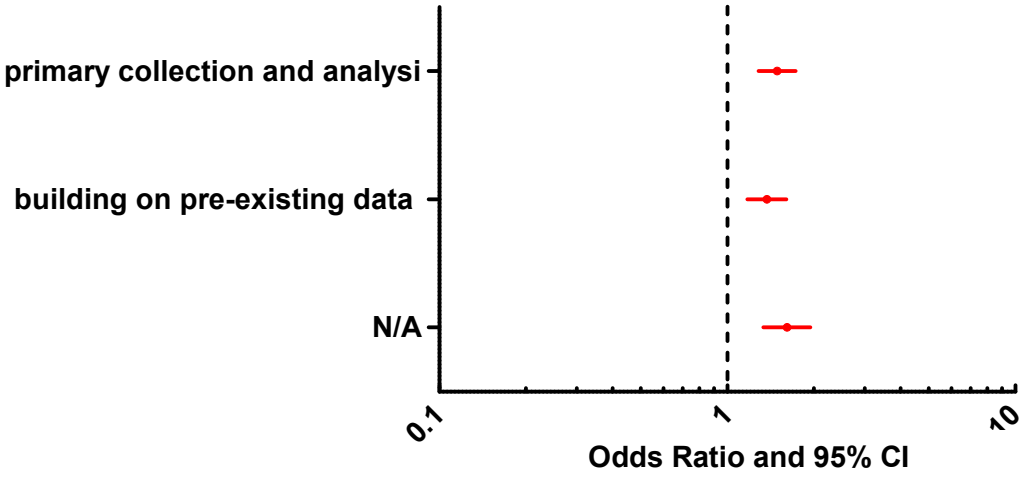
Were case control groups related?



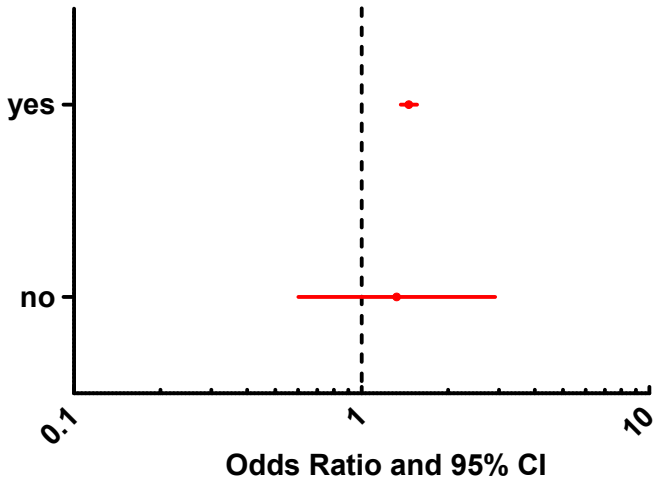
Were cases/controls drawn from the same population?



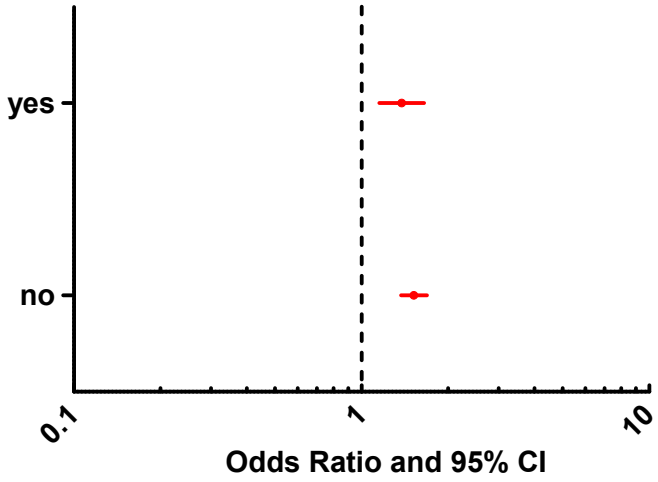
Was the data source:



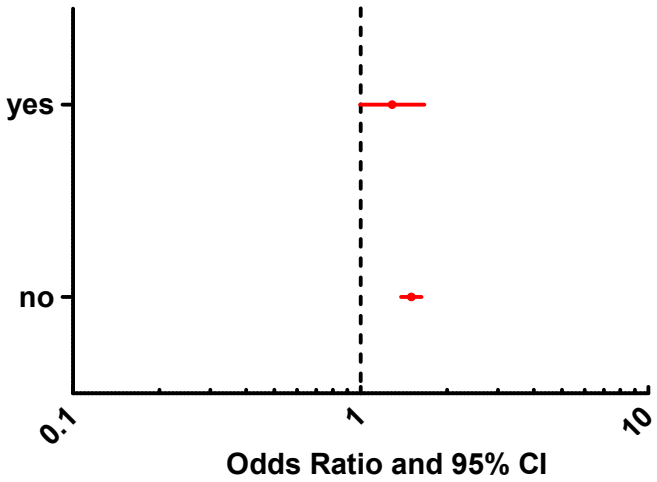
Was quality assurance reported?



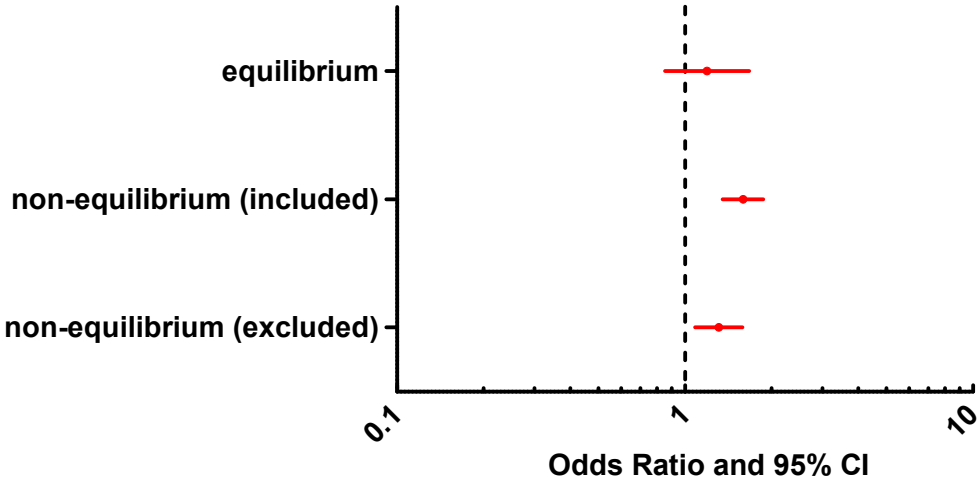
Did the study report power/sample size calculations?



Was deviation from HWE tested?



If so, what was the result?

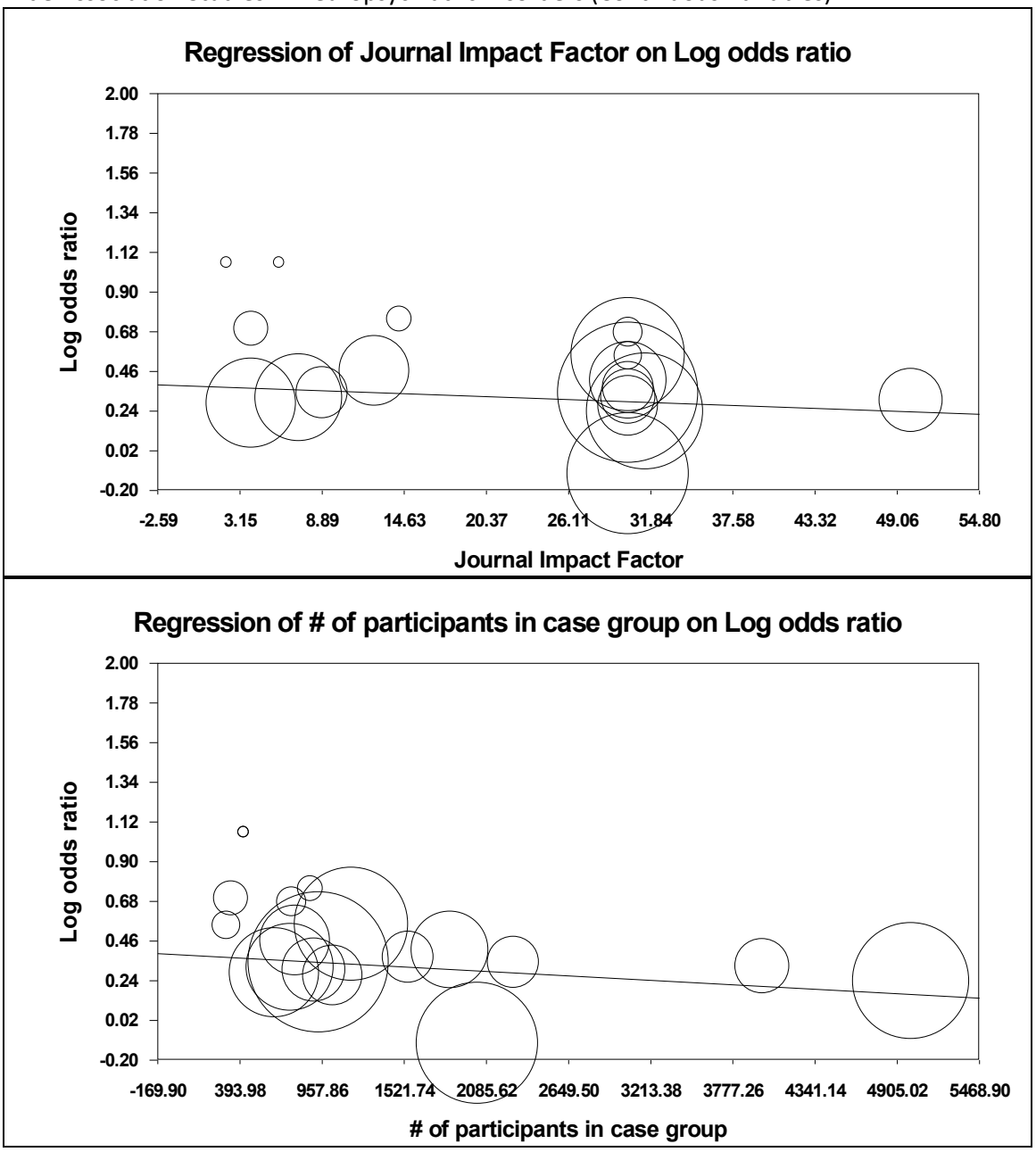


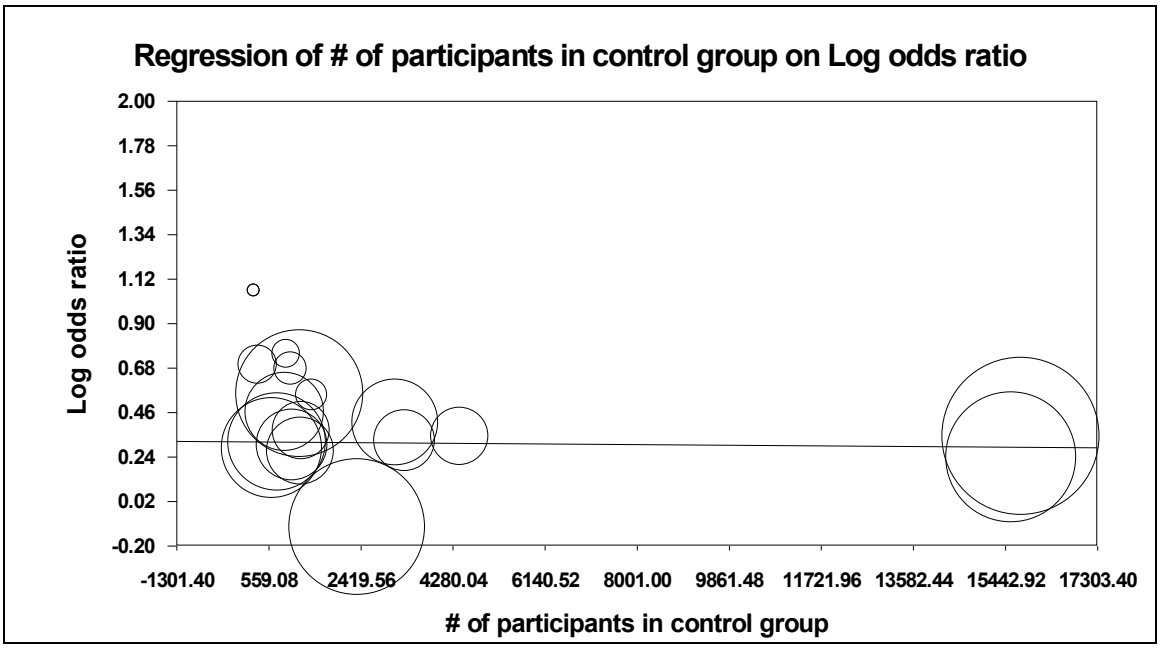
Univariate meta-regression under a random effects model using unrestricted maximum likelihood, adjusting for journal impact factor, generated a between-study heterogeneity Q of 1.40 (p = 0.236) (Table 23 and Figure 13). Adjusting for the number of participants recruited for the case group, the between-study heterogeneity Q was 2.99 (p = 0.084). Reporting the number of participants in the control group generated a between-study heterogeneity Q of 0.97 (p = 0.324).

Table 23: Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables)

Methodological Characteristic	*Slope*	*95% CI*	*Q-value (p-value)*
Journal Impact Factor	-0.00284	(-0.00602, 0.00034)	1.40 (0.236)
Number of study participants in case group	-0.00004	(-0.00007, -0.00002)	2.99 (0.084)
Number of study participants in control group	<-0.00000	(-0.00001, <0.00000)	0.97 (0.324)

Figure 10: Scatter Plots of Results of Meta-Regression Analysis of Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders (Continuous Variables)

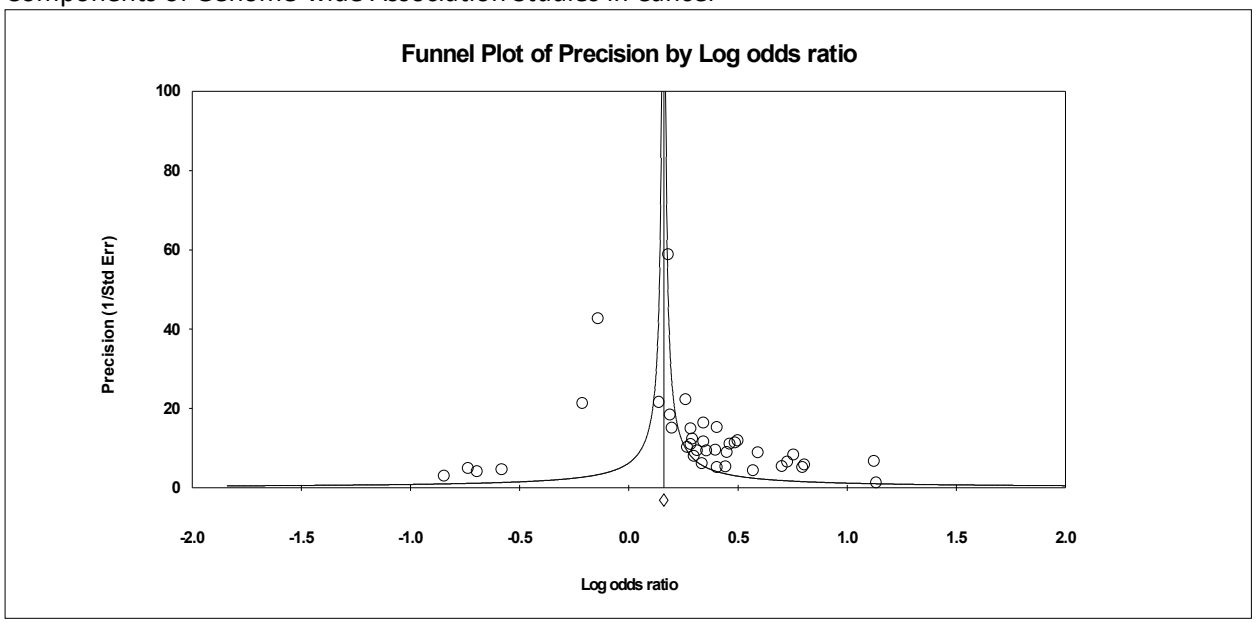



Regression of # of participants in control group on Log odds ratio
Log odds ratio
2.00
1.78
1.56
1.34
1.12
0.90
0.68
0.46
0.24
0.02
-0.20
-1301.40
559.08
2419.56
4280.04
6140.52
8001.00
9861.48
11721.96
13582.44
15442.92
17303.40
# of participants in control group

2.5 Study Size Effects

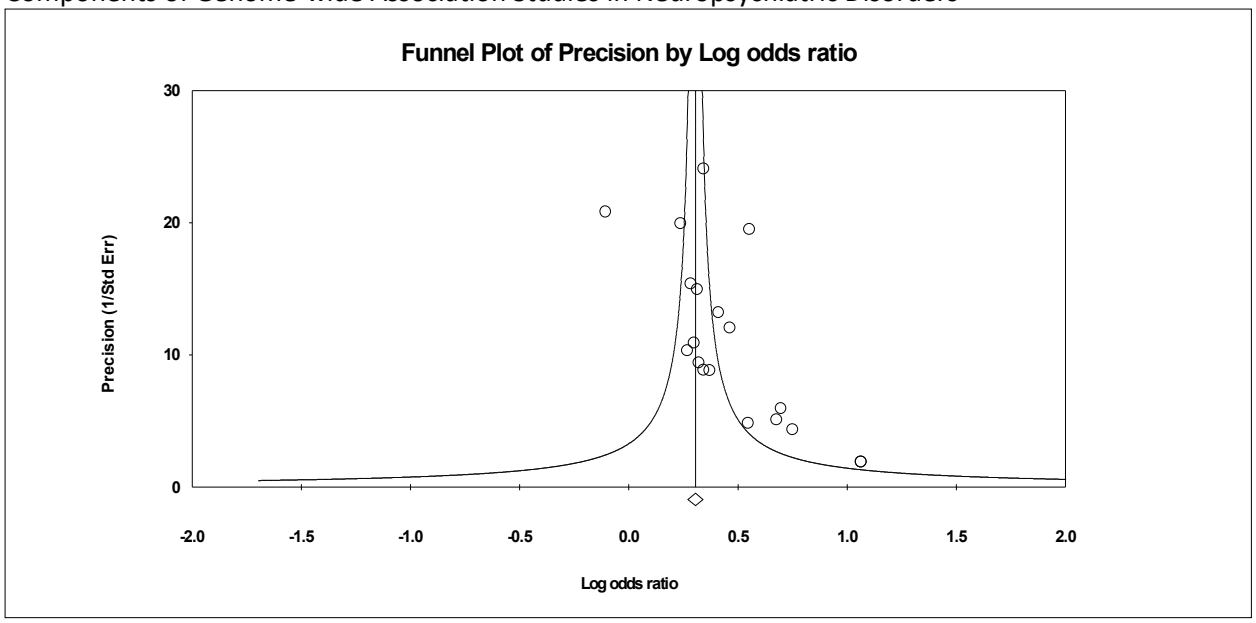
The funnel plot of precision by log odds ratio of the replication components of genome-wide association studies in cancer is shown in Figure 11. The Classic fail-safe N value for this set of studies was 3278, implying that 3278 missing studies (86.26 missing studies for every 1 included study) would need to be located and included in the analysis in order for the observed effect of publication bias to be nullified, demonstrating that publication bias exists in this set of studies. Egger's regression intercept was 2.37 (95% CI 0.69, 4.06).

Figure 11: Assessment of Publication Bias by Funnel Plot of Precision by Log odds ratio of the Replication Components of Genome-wide Association Studies in Cancer



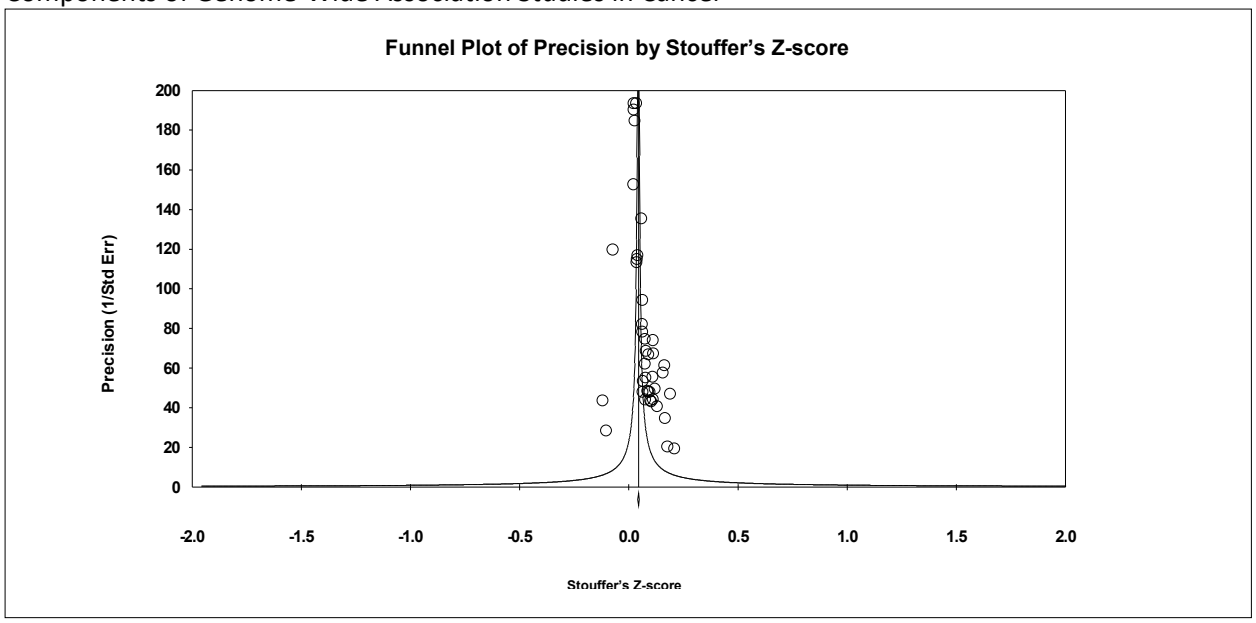
The funnel plot of precision by log odds ratio of the replication components of genome-wide association studies in neuropsychiatric disorders is shown in Figure 14. The Classic fail-safe N value for this set of studies was 1432, implying that 1432 missing studies (75.37 missing studies for every 1 included study) would need to be located and included in the analysis in order for the observed effect of publication bias to be nullified, demonstrating that publication bias exists in this set of studies. Egger's regression intercept was 2.11 (95% CI -0.28, 4.50).

Figure 12: Assessment of Publication Bias by Funnel Plot of Precision by Log odds ratio of the Replication Components of Genome-wide Association Studies in Neuropsychiatric Disorders



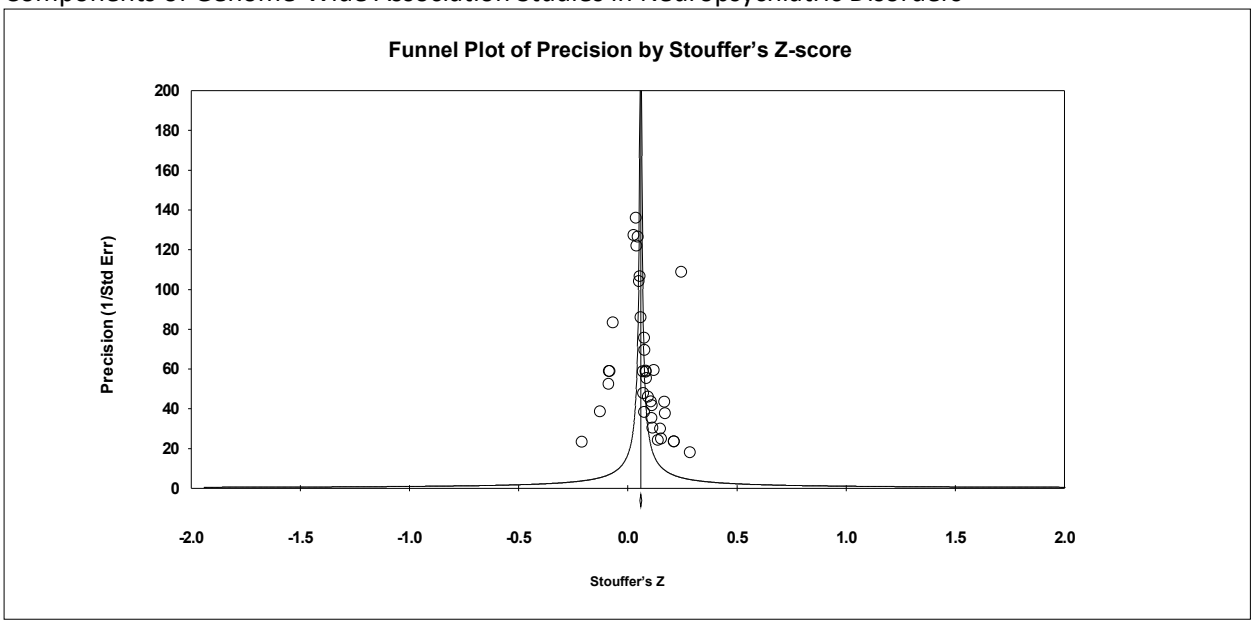
The funnel plot of precision by log odds ratio of the index components of genome-wide association studies in cancer is shown in Figure 4. The Classic fail-safe N value for this set of studies was 8953, implying that 8953 missing studies (223.83 missing studies for every 1 included study) would need to be located and included in the analysis in order for the observed effect of publication bias to be nullified, demonstrating that publication bias exists in this set of studies. Egger's regression intercept was 4.33 (95% CI 2.23, 6.43).

Figure 13: Assessment of Publication Bias by Funnel Plot of Precision by Stouffer's Z-score of the Index Components of Genome-Wide Association Studies in Cancer



The funnel plot of precision by log odds ratio of the index components of genome-wide association studies in neuropsychiatric disorders is shown in Figure 8. The Classic fail-safe N value for this set of studies was 5073, implying that 5073 missing studies (140.92 missing studies for every 1 included study) would need to be located and included in the analysis in order for the observed effect of publication bias to be nullified, demonstrating that publication bias exists in this set of studies. Egger's regression intercept was 1.17 (95% CI -2.57, 4.91).

Figure 14: Assessment of Publication Bias by Funnel Plot of Precision by Stouffer's Z-score of the Index Components of Genome-Wide Association Studies in Neuropsychiatric Disorders



A summary of the results of the tests for publication bias is shown in Table 24.

Table 24: Results of Tests for Publication Bias in Index and Replication Components of Genome-wide Association Studies in Cancer and Neuropsychiatric Disorders

Group	*Classic fail-safe N (raw value)*	*Classic fail-safe N (rate/study)*	*Egger's regression intercept*
Index (Cancer)	8953	223.83	4.33 (2.23, 6.43)
Index (Neuropsychiatric Disorders)	5073	140.92	1.17 (-2.57, 4.91)
Replication (Cancer)	3278	86.26	2.37 (0.69, 4.06)
Replication (Neuropsychiatric Disoders)	1432	75.37	2.11 (-0.28, 4.50)

DISCUSSION

Information collected from the independent samples of case-control studies investigating gene-disease associations identified by Yesupriya et al[38] (years of publication: 2001-2003 and 2006) and Aljasir[39] (2007) was compared to identify if any changes in reporting practices of genetic association studies had occurred over time. The studies published in 2007 were significantly more likely to state that this was the first study on the specific issue, use a larger group of study participants, report the available power of the study, report that all genetic variants were examined for Hardy-Weinberg equilibrium, and identify that any polymorphism failed Hardy-Weinberg equilibrium given that Hardy-Weinberg equilibrium was examined. In contrast to this, the studies published in 2007 were significantly less likely to report clear information regarding the population from which cases and controls were drawn, and explicitly state the use of unrelated study participants. Given this information, we can identify a few trends in the reporting of case-control studies investigating gene-disease associations.

Firstly, human genome epidemiology research has been in methodological flux, initially with a shift in emphasis from family-based studies to case-control studies with unrelated controls, and later from studies of candidate genes to genome-wide association studies.[222,223] Secondly, larger studies are being undertaken and the reporting of statistical analysis has become more transparent. Thus, power calculations and the methods and results of testing for Hardy-Weinberg equilibrium are increasingly being reported. On the other hand, however, certain qualitative characteristics of the studies, such as population characteristics and documentation of inclusion or exclusion of related participants, have become less transparently reported.

A characteristic of the findings of genome-wide association studies is that, in general, the effect sizes of replicated genetic associations for common diseases and traits have been small and account for only a small proportion of estimated heritabilities.[224] This has led to the common disease-common

variant hypothesis being challenged,[225-228] with increasing interest in the effects of rare variants, types of genetic variation other than SNPs and the effects of possible gene-environment and gene-gene interactions. Furthermore, there have been advances in technologies enabling the capture of large amounts of data on types of biologic variation in addition to genomic variation (including epigenetic, proteomic, transcriptomic and metabolomic). Therefore, considerable methodological flux in the field of human genome epidemiology is likely to continue.

In the conclusion of their paper, Yesupriya et al[38] expressed concerns regarding inadequate reporting, in particular that many details required to assess the validity of the study findings were not consistently reported. Subsequently, the present comparison between the work performed by Yesupriya et al[38] and Aljasir[39] identified an improvement in the reporting of certain characteristics. Although biased outcomes have been found to be associated with poor reporting of findings in genetic association studies,[39,193-196,229] the reporting of information within case-control studies is moving towards greater transparency, allowing readers to more easily appreciate and assess the work being done. These ideas formed the basis upon which this research was performed, and allowed the current investigator to evaluate the transparency of reporting in genome-wide association studies.

During the development of the protocol for this research, there was extensive discussion between the current investigator and advisory committee on how to best capture a suitable outcome in the meta-analysis of the index components. Various suggestions were made, including creating a pattern of the measure of uncertainty (p-value) through the use of q-q plots, extracting the median measure of uncertainty (p-value) or extracting the number of SNPs used for discovery. Ultimately it was decided that, given the challenge that these options were likely to present due to the limits of reporting, extracting the most extreme reported measure of uncertainty (p-value) would be chosen as the most suitable outcome measure in the index components of the genome-wide association studies.

Now that the research has been completed, there is a debate to be had regarding whether or not the analysis plan that was originally developed could be improved upon. Rather than reporting the odds ratio for the replication components in the abstract of each journal article, the majority of authors combined the data from the index and replication components and chose to report a combined odds ratio. With this knowledge, future studies of this nature could be improved by utilizing the most extreme combined odds ratio as the primary outcome measure, rather than the most extreme reported measure of uncertainty (p-value) from the index components of the study and odds ratio from the replication components of the study. This would ensure that only validated associations would be taken forward to future stages of research. That being said, an approach of this nature could arguably increase the amount of false negatives a genome-wide association study would produce.[240] At any rate, it is clear that incorporating the evidence from this research with previous findings about genome-wide association studies must be done if future research of this nature is to be carried out.

In the index components, out of fifty-three studies on cancer and seventy-four studies on neuropsychiatric disorders, two (3.8%) were excluded from the group on cancer and two (2.7%) were excluded from the group on neuropsychiatric disorders due to lack of reporting of the p-value. In addition to this, 11 (20.8%) were excluded from the group on cancer and 36 (48.6%) were excluded from the group on neuropsychiatric disorders due to lack of reporting of the direction of effect. Omitting journal articles on the basis of a lack of reporting of the direction of effect likely had a considerable effect on the meta-analysis. The measures of uncertainty (p-values) excluded from further analysis on the basis of this exclusion criterion range from $1.90e^{-3}$ to $3.00e^{-12}$ of studies on cancer and from $1.50e^{-2}$ to $2.30e^{-44}$ of studies on neuropsychiatric disorders. In comparison to this, the measures of uncertainty (p-values) included in further analyses range from $6.22e^{-2}$ to $8.00e^{-24}$ of studies on cancer and from $5.00e^{-2}$ to $1.80e^{-157}$ of studies in neuropsychiatric disorders. Without a clear indication of the direction of effect of the values extracted from the excluded groups of journal articles, it is difficult to hypothesize

how the meta-analysis would be affected. Additionally, any post-hoc analysis was carefully avoided in order to retain the integrity of the original analysis plan of this research. In future studies of this nature, consideration could be made as to the most appropriate statistical method chosen to perform the meta-analysis on the index components of genome-wide association studies. In addition to this, it would be pertinent for the authorship at large to identify this lack of transparency of reporting of the direction of effect of the reported measure of uncertainty (p-value) as an important way to improve the reporting of the index components of future genome-wide association studies.

A small group of studies (three (6.7%) of forty-five studies from the group on cancer and three (7.7%) of thirty-nine studies from the group on neuropsychiatric disorders) reported asymmetrical confidence limits for the odds ratios in the replication component of the genome-wide association studies. These studies were excluded from the meta-analysis as it was not possible to incorporate the asymmetrical confidence limits into the statistical techniques employed. As the scale of odds ratios is asymmetric, and only becomes symmetric following a log transformation of their true values,[234] the most likely explanation for these asymmetrical confidence limits is that the investigators of the studies that reported odds ratios with asymmetric confidence limits decided to use the true form of their odds ratios, disregarding the opportunity to log transform their odds ratios for the purposes of symmetry. In future studies of this nature, consideration could be made as to the most appropriate statistical method chosen to perform the meta-analysis on the replication components of genome-wide association studies. In addition to this, it would be pertinent for the authorship at large to identify this lack of transparency of reporting of the log-transformed odds ratios as an important way to improve the reporting of the replication components of future genome-wide association studies.

Regarding the index components of each group of studies, 32.1% of studies on cancer included at least 5,000 study participants, while only 17.6% of studies on neuropsychiatric disorders included at

least 5,000 study participants. This study confirmed the findings presented by Manolio et al[31] by demonstrating that the index components of the studies on cancer had larger sample sizes than the index components of the studies on neuropsychiatric disorders. There was also considerable heterogeneity between the studies, as demonstrated by a Q of 659.35 ($p < 0.0001$) and an I^2 statistic of 94.1 for the studies on cancer, and a Q of 983.05 ($p < 0.0001$) and an I^2 statistic of 96.4 for the studies on neuropsychiatric disorders. Out of all of the studies, only nine studies demonstrated a Stouffer's Z-score with a negative direction of effect (three out of forty studies on cancer and six out of thirty-six studies on neuropsychiatric disorders).

Regarding the replication components of each group of studies, 57.8% of the studies on cancer included at least 5,000 study participants and showed a summary odds ratio of 1.34 (95% CI 1.25, 1.43), while only 28.2% of the studies on neuropsychiatric disorders included at least 5,000 study participants and showed a summary odds ratio of 1.43 (95% CI 1.30, 1.57). Our study confirmed the findings presented by Manolio et al[31] by demonstrating that the replication components of the studies on cancer had larger sample sizes and smaller odds ratios than the replication components of the studies on neuropsychiatric disorders. There was also considerable heterogeneity between the studies, as demonstrated by a Q of 595.84 ($p < 0.0001$) and an I^2 statistic of 93.3 for the studies on cancer, and a Q of 145.52 ($p < 0.0001$) and an I^2 statistic of 86.9 for the studies on neuropsychiatric disorders. Out of all of the studies, only four studies demonstrated an odds ratio with a negative direction of effect (three out of forty-one studies on cancer and one out of nineteen studies on neuropsychiatric disorders).

Both groups of genome-wide association studies tended to be published in a select few journals. The majority of each group of studies was published in *Nature Genetics*: thirty-one of forty-one studies on cancer and twelve of thirty-six studies on neuropsychiatric diseases. In all cases, studies published in

Nature Genetics tended to demonstrate low Stouffer's Z-scores and odds ratios with narrow confidence limits.

In both groups of studies, those of Japanese origin tended to demonstrate higher Stouffer's Z-scores when compared to the other countries of origin, while the magnitude of odds ratios tended to be similar regardless of country of origin. There was a notable absence of publications from China, a country second only to the United States of America in terms of the number of journal articles published.[197,230] This may be because in the time period under review the cost of genotyping with genome-wide chips had not reduced sufficiently to make such studies feasible during the time period in this country, because of a focus on other types of genetic association studies. It is more likely, however, that these studies were published in languages other than English.

Similar to type of journal, there was a diverse collection of types of diseases investigated in both studies on cancer and studies on neuropsychiatric disorders, making it difficult to infer any trends in the analysis of the index components based on type of disease. In the index components, Stouffer's Z-scores with negative correlations were observed with ovarian cancer, attention deficit/hyperactivity disorder, autism and major depressive disorder. In the replication components, odds ratios with negative correlations were observed with follicular lymphoma and glioma. Stratifying the data by disease presents some curious trends, however the small sample sizes prevented the current investigator from making more powerful conclusions based on the data.

The magnitude of the Stouffer's Z-scores and odds ratios (significant with narrow, positive confidence limits) tended to be similar in both studies on cancer and studies on neuropsychiatric disorders according to the type of control group (population-based vs. hospital-based vs. mix), whether or not case and control groups included related participants, whether or not cases and controls were reported to be drawn from the same source population, if the data were newly collected specifically for

the study reported or built on pre-existing data, and whether or not the study reported power or sample size calculations.

Studies that reported performing a Hardy-Weinberg calculation in the index portion of their genome-wide association study demonstrated Stouffer's Z-scores that were significant positive values with narrow confidence limits. Contrary to this, studies that did not report Hardy-Weinberg calculations demonstrated Stouffer's Z-scores that were statistically insignificant. According to this result, studies that employ the use of Hardy-Weinberg equilibrium in their index components are properly addressing quality control at the level of the single-nucleotide polymorphisms (SNPs) being studied, ensuring the quality of the replication components in the genome-wide association study.

Stratifying the results for the reporting of Hardy-Weinberg calculations in the replication components gave interesting results. In both sets of studies, those that reported testing Hardy-Weinberg equilibrium demonstrated summary odds ratios that were insignificant, while those that did not report testing Hardy-Weinberg equilibrium demonstrated summary odds ratios that were significant with positive correlations. This is an important result, as the make-up of the population can be questioned in those studies that did not report Hardy-Weinberg calculations, leading to further scrutiny regarding their study sample selection methods. Without further investigation into these studies, it is impossible to determine whether the groups publishing the studies that did not report Hardy-Weinberg equilibrium actually performed it and simply refrained from reporting the result, or if the groups simply failed to perform Hardy-Weinberg calculations at all. Future studies could improve the scope of the current work if reporting the use of Hardy-Weinburg equilibrium is made a priority.

In the index components of genome-wide association studies, there were contrasting results for the trends found regarding journal impact factor. For the studies on cancer, Stouffer's Z-scores tended to decrease as the magnitude of the journal impact factor increased, suggesting that journals that have

a higher impact publish studies that identify less significant associations. Contrary to this, however, for the studies on neuropsychiatric disorders, Stouffer's Z-scores tended to increase as the magnitude of the journal impact factor increased, suggesting that journals that have a higher impact publish studies that identify more significant associations. In the replication components of genome wide association studies, odds ratios tended to decrease as the magnitude of the journal impact factor increased for both studies on cancer and studies on neuropsychiatric disorders. Despite these results, since, as shown above, a select few journals published the vast majority of the sampled journal articles, this trend may be attributed to a lack of sample size (or, in the case of *Nature* Genetics, a high journal impact factor) rather than any statistically meaningful association.

Supplementary materials, in online format, were included with the majority of journal articles investigating both cancer and neuropsychiatric disorders. It is the opinion of the current investigator than improving the format by which supplementary materials are included with the journal article must be made a priority. Sifting through the multitude of information included with each journal article proved rather difficult and time consuming, as most supplementary material packages were devoid of a rational, organized structure. In the same way that journals provide prospective authors with instructions on manuscript preparation and submission requirements, so to should they provide a statement regarding not only how to properly structure, but also limit the supplementary material they intend to include with their journal article.

Emilie Marcus has identified that supplementary material presents a significant advantage in the online age by allowing extra information to be presented outside the confines of the published journal article.[241] Despite this, she claims it has become a place where authors, restricted by the limits presented to them by scientific journals, can place the results that "don't fit" into the main body of the article.[241] In fact, one of the journal articles included in the current study made reference to a link

containing 443 pages of supplementary material.[189] In light of this trend, *Cell* has taken it upon itself to develop a new way of online publishing. Their "Article of the Future" will include new and innovative navigational techniques, incorporating an integrated format which will include in-depth information for all sections of each journal article, rather than the traditional method of attaching supplemental information after the journal article.[242] This concept, although still in its infancy, has motivated publishers to think about how to improve their online format in an ever-changing scientific world.

Not surprisingly, in view of the design of genome-wide association studies, which minimize false positive results, and the selection of the most extreme measures of uncertainty (p-values) and odds ratios as outcome measures, asymmetrical funnel plots produced results that would be compatible with bias towards selective publishing of results of genome-wide association studies that report a positive direction of effect. A positive direction of effect was seen predominantly in the group of journal articles on cancer, the group of studies that tended to have larger sample sizes and smaller odds ratios. In comparison, the group of journal articles on neuropsychiatric disorders, those with smaller sample sizes and larger odds ratios, demonstrated weaker results towards the selective publishing of results with a positive direction of effect. Since the group of journal articles regarding cancer demonstrated larger overall study sizes, there seems to be an effect not only on the direction of effect, but also on study size. Although not explicitly extracted from the journal articles, it is likely that the investigators may have chosen reference groups that would produce this pattern of results, mirroring past trends that demonstrate the vast majority of published journal articles consisting of significant positive results, rather than negative or neutral results.[198-202] As most genome-wide association studies included a replication component, these results could also be interpreted as indicating that the gene-disease associations published do indeed consist more often than not of true positive associations.

It was the intent of the current investigator, upon completion of this research, to use the information to assist in drafting an adjunct to the STrengthening the REporting of Genetic Association studies (STREGA) Statement, so that the efforts made by those responsible for creating guidelines to assist in the reporting of other types of genetic association studies would be mirrored by those attempting to report on genome-wide association studies. The STREGA Statement has proven a useful tool for improving the reporting of results of genetic association studies. In order to improve the reporting of genome-wide association studies, items specific to the methods (e.g., source and composition of study participants, sample of SNPs analyzed, inclusion of Q-Q plots, utilization of log-transformed odds ratios) and results (e.g., directions of effect of measures of uncertainty (p-values), summary odds ratios and confidence limits) of genome-wide association studies need to be considered for inclusion. Unfortunately, owing to the lack of a significant level of heterogeneity shown within each group of studies, specifically regarding their methodological characteristics (in particular within the replication studies), it was concluded that, on the basis of empirical evidence, it would be difficult to draft a proper extension to the STREGA Statement for genome-wide association studies. As the current investigator has also identified the role that the effect of small individual studies has played within and between each group of studies, in order to draft a proper extension to the STREGA Statement for genome-wide association studies it would be necessary to identify a sample of studies that better represents the overall collection of published journal articles. Using a more representative, random sample of journal articles on genome-wide association studies would be required.

CONCLUSION

Following the completion of the Human Genome Project in 2003 and the International HapMap Project in 2005, the utilization of genome-wide association studies became possible. Since then, there has been an explosion in the production and publication of this approach in the scientific literature, as the impact genome-wide association studies could have on the field of medicine has been postulated to be great.[220] To date, and to the best of the investigator's knowledge, a review of the reporting characteristics of genome-wide association studies has not been performed. Therefore, the purpose of this research was to perform a large-scale analysis of genome-wide association studies in order to produce evidence useful in guiding their reporting and synthesis.

An analysis of information collected from three independent samples of case-control studies investigating gene-disease associations from the years 2001-2003, 2006 and 2007 was performed in order to identify if any changes in the results of the extracted information had occurred over time, in order to create a link to the information collected from the sample of genome-wide association studies. Yesupriya et al[38] concluded that the methodological characteristics they deemed necessary in assessing error and bias in gene-disease association studies were not consistently reported, and needed the help of reporting guidelines if improvements were going to be seen. Following this, Aljasir[39] reported evidence which continued to suggest that these methodological characteristics are associated with the magnitudes of effect observed, and efforts to improve the reporting of studies needed to continue.

In consideration of these efforts, this research was able to demonstrate that although the transparency of reporting has improved over time, efforts still need to be made in order to improve the process. As the methodological characteristics in genome-wide association studies have been demonstrated to be associated with the magnitudes of effect observed (much like in the work done by

both Yesupriya et al[38] and Aljasir[39]), this work can be used to guide future work being done to improve the reporting of human genome epidemiology association studies.

As this was, to the best of the current investigator's knowledge, the first large-scale study performed to identify trends in the reporting of methodological characteristics within a set of genome-wide association studies, it can be used as a launching point for future endeavours on the subject. Certain methods used in this research were exposed for their strengths and weaknesses, which can be addressed with further work. For example, future studies could incorporate groups of studies with larger sample sizes, identify a more suitable way to analyze the outcome measure (e.g., utilizing log-transformed odds ratios in order to avoid asymmetric confidence limits), or include the reporting characteristics identified by this research as useful in the attempt to improve the reporting of genome-wide association studies.

This research attempted to provide evidence that the characteristics of genome-wide association studies are being transparently reported. It has been found that, despite reasonably sound reporting practices, there are still many inconsistencies in the reported information, and the reporting of genome-wide association studies requires improvement in order to achieve true transparency. The production and eventual usage of a set of reporting guidelines specific to genome-wide association studies will vastly improve this process, and will provide authors the chance to identify what an acceptable genome-wide association study should include. As the usage of genome-wide association studies continues to increase,[243] it is vital that the overall reporting of genome-wide association studies be improved, so as to benefit the entire scientific community.

REFERENCES

1. Little J, Khoury MJ, Bradley L, Clyne M, Gwinn M, Lin B, Lindegren M-L, Yoon P. The human genome project is complete. How do we develop a handle for the pump? American Journal of Epidemiology 2003;157(8):667-673.
2. Khoury MJ, Millikan R, Little J, Gwinn M. The emergence of epidemiology in the genomics age. International Journal of Epidemiology 2004;33(5):936-944.
3. Genomics, Health and Society Working Group. *Genomics, Health and Society. Emerging Issues for Public Policy.* Ottawa: Government of Canada Policy Research Initiative 2004. Available at http://www.policyresearch.gc.ca/doclib/genomicbook_e.pdf. Accessed October 30, 2008.
4. Hindorff LA, Junkins HA, Mehta JP, Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at http://www.genome.gov/26525384. Accessed November 8, 2008.
5. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Yong Zou G, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N. STrengthening the REporting of Genetic Association Studies (STREGA) – An Extension of the STROBE Statement. Available at http://www.medicine.uottawa.ca/public-health-genomics/web/assets/documents/Final-STREGA%20manuscript-Feb2%202009.pdf. Accessed February 27, 2009.
6. The EQUATOR Network. *Enhancing the QUAlity and Transparency Of health Research.* The EQUATOR Network 2008. Available at http://www.equator-network.org/index.aspx?o=1032. Accessed November 9, 2008.

7. Schulz KF, Altman D, Moher D, CONSORT Group. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. Journal of Clinical Epidemiology 2010;63(8):834-840.

8. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. Annals of Internal Medicine 2007;147(8):573-577.7

9. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Public Library of Science: Medicine 2009;6(7):e1000097.

10. Ioannidis JPA, Gwinn M, Little J, Higgins JPT, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin M-R, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khouly MJ. Human Genome Epidemiology Network and the Network of Investigator Networks. A road map for efficient and reliable human genome epidemiology. Nature Genetics 2006;38(1):3-5.

11. Knoppers BM, Fortier I, Legault D, Burton P. The Public Population Project in Genomics (P^3G): a proof of concept? European Journal of Human Genetics 2008;16(6):664-665.

12. National Institutes of Health: . *Policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS).* Federal Register 2007;72(166):49290-49297. Available at http://www.grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html . Accessed November 8, 2008.

13. Christensen K, Murray JC. What genome-wide association studies can do for medicine. New England Journal of Medicine 2007;356(11):1094-1097.

14. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex studies. Nature Reviews: Genetics 2005;6(2):95-108.
15. Hunter DJ, Kraft P. Drinking from the fire hose—statistical issues in genomewide association studies. New England Journal of Medicine 2007;357(5):436-439.
16. Little J, Bradley L, Bray MS, Clyne M, Dorman J, Ellsworth DL, Hanson J, Khoury 7M, Lau J, O'Brien TR, Rothman N, Stroup D, Taioli E, Thomas D, Vainio H, Wacholder S, Weinberg C. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. American Journal of Epidemiology 2002;156(4):300-310.
17. Fremier NB, Sabatti C. Human genetics: Variants in common disease. Nature 2007;445(7130):828-830.
18. Nicholson JK, Holmes E, Elliott P. The Metabolome-Wide Association Study: A New Look at Human Disease Risk Factors. Journal of Proteome Research 2008;7(9):3637-3638.
19. Budarf ML, Labbe C, David G, Rioux JD. GWA studies: rewriting the story of IBD. Trends in Genetics 2009;25(3):137-146.
20. Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nature Genetics 2001;29(3):306-309.
21. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genetics in Medicine 2002;4(2):45-61.
22. Gorroochurn P, Hodge SE, Heiman GA, Durner M, Greenberg DA. Non-replication of association studies: "pseudo-failures" to replicate? Genetics in Medicine 2007;9(6):325-331.
23. Von Elm E, Egger M. The scandal of poor epidemiological research. British Medical Journal 2004;329(7471):868-869.
24. Altman DG, Moher D. Developing guidelines for reporting healthcare research: scientific rationale and procedures. Medicina Clinica 2005;125(Supplement 1):8-13.

25. Gluud LL. Bias in Clinical Intervention Research. American Journal of Epidemiology 2006;163(6):493-501.
26. Pearson TA, Manolio TA. How to Interpret a Genome-wide Association Study. Journal of the American Medical Association 2008;299(11):1335-1344.
27. International HapMap Project. *About the HapMap*. International HapMap Project 2008. Available at http://www.hapmap.org/thehapmap.html.en. Accessed December 5, 2008.
28. Johnson AD, O'Donnell AJ. An Open Access Database of Genome-wide Association Results. BMC Medical Genetics 2009;10(6).
29. PubMed. *PubMed Literature Search*. National Center for Biotechnology Information 2008. Available at http://www.ncbi.nlm.nih.gov/pubmed/. Accessed December 18, 2008.
30. Human Genome Epidemiology Network. *HuGE Navigator*. Human Genome Epidemiology Network 2008. Available at http://www.hugenavigator.net/. Accessed December 18, 2008.
31. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. The Journal of Clinical Investigation 2008;118(5):1590-1605.
32. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make of the results of meta-analyses? Journal of Clinical Epidemiology 2000;53(9):964-972.
33. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. Journal of Clinical Epidemiology 2006;58(8):769-776.
34. Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JPA. Local Literature Bias in Genetic Epidemiology: An Empirical Evaluation of the Chinese Literature. Public Library of Science: Medicine 2005;2(12):1309-1317.

35. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. British Medical Journal 1997;315(7109):629-634.
36. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? Journal of Clinical Epidemiology 2000;53(9):964-972.
37. Pham B, Klassen TP, Lawson ML, Moher D. Language of publication restrictions in systematic reviews gave different results depending on whether the intervention was conventional or complementary. Journal of Clinical Epidemiology 2005;58(8):769-776.
38. Yesupriya A, Evangelou E, Kavvoura FK, Patsopoulos NA, Clyne M, Walsh MC, Lin BK, Yu W, Gwinn M, Ioannidis JPA, Khoury MJ. Reporting of Human Genome Epidemiology (HuGE) association studies: An empirical assessment. BMC Medical Research Methodology 2008;8:31.
39. Aljasir B. (2009). *Impact of methodological characteristics of case-control studies investigating gene-disease associations on knowledge outcomes*. Master's Thesis. University of Ottawa: Canada.
40. Freely Associating. Nature Genetics 1999;22(1):1-2.
41. Gambano G, Anglani F, D'Angelo A. Association of studies of genetic polymorphisms and complex disease. The Lancet 2000;355(9200):308-311.
42. Cardon LR, Bell JI. Association study designs for complex diseases. Nature Reviews: Genetics 2001;2(2):91-99.
43. Tabor HK, Risch NJ, Myers RM. OPINION: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews: Genetics 2002;3(5):391-397.
44. In search of genetic precision. The Lancet 2003;361(9355):357.

45. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics 2003;33(2):177-182.

46. Taubes G. Epidemiology faces its limits. Science 1995;269(5221):164-168.

47. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. Journal of the American Medical Association 2000;283(15):2008-2012.

48. Ioannidis JPA, Bernstein J, Boffeta P, Danesh J, Dolan S, Hartge P, Hunter D, Inskip P, Jarvelin MR, Little J, Maraganore DM, Newton Bishop JA, O'Brien TR, Petersen G, Riboli E, Seminara D, Taioli E, Uitterlinden AG, Vineis P, Winn DM, Salanti G, Higgins JPT, Khoury MJ. A Network of Investigator Networks in Human Genome Epidemiology. American Journal of Epidemiology 2005 162(4):302-304.

49. Lin BK, Clyne M, Walsh M, Gomez O, Yu W, Gwinn M, Khoury MJ. Tracing the Epidemiology of Human Genes in the Literature: The HuGE Published Literature Database. American Journal of Epidemiology 2006;164(1):1-4.

50. Pinto D, Marshall C, Feuk L, Scherer SW. Copy-number variation in control population cohorts. Human Molecular Genetics 2007;16:R168-R173.

51. Bierut LJ, Madden PAF, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG. Novel genes identified in a high-density genome wide association study for nicotine dependence. Human Molecular Genetics 2007;16(1):24-35.

52. Hillmer AM, Brockschmidt FF, Hanneken S, Eigelshoven S, Steffens M, Flaquer A, Herms S, Becker T, Kortüm A-K, Nyholt DR, Zhao ZZ, Montgomery GW, Martin NG, Mühleisen TW, Alblas

MA, Moebus S, Jöckel K-H, Bröcker-Preuss M, Erbel R, Reinartz R, Betz RC, Cichon S, Propping P, Baur MP, Wienker TF, Kruse R, Nöthen MM. Susceptibility variants for male-pattern baldness on chromosome 20p11. Nature Genetics 2008;40(11):1279-1281.

53. Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE, Downes K, Barrett JC, Healy BC, Mychaleckyj JC, Warram JH, Todd JA. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nature Genetics 2008;40(12):1399-1401.
54. Suzuki S, Yoshimura M, Nakayama M, Abe K, Yamamuro M, Nagayoshi Y, Kojima S, Kaikita K, Sugiyama S, Yasue H, Ogawa H. A novel genetic marker for coronary spasm in women from a genome-wide single nucleotide polymorphism analysis. Pharmacogenetics and Genomics 2007;17(11):919-930.
55. Gu YY, Harley ITW, Henderson LB, Aronow BJ, Vietor I, Huber LA, Harley JB, Kilpatrick JR, Langefeld CD, Williams AH, Jegga AG, Chen J, Wills-Karp M, Arshad SH, Ewart SL, Thio CL, Flick LM, Filippi M-D, Grimes HL, Drumm ML, Cutting GR, Knowles MR, Karp CL. Identification of *IFRD1* as a modifier gene for cystic fibrosis lung disease. Nature 2009;458(7241):1039-1042.
56. Zemunik T, Boban M, Lauc G, Janković S, Rotim K, Vatavuk Z, Benčić G, Đogaš Z, Boraska V, Torlak V, Sušac J, Zobić I, Rudan D, Pulanić D, Modun D, Mudnić I, Gunjaća G, Budimir D, Hayward C, Vitart V, Wright AF, Campbell H, Rudan I. Genome-wide Association Study of Biochemical Traits in Korčula Island, Croatia. Genomics 2009;50(1);23-33.
57. Capon F, Bijlmakers M-J, Wolf N, Quaranta M, Huffmeier U, Allen M, Timms K, Abkevich V, Gutin A, Smith R, Warren RB, Young HS, Worthington J, Burden AD, Griffiths CEM, Hayday A, Nestle FO, Reis A, Lanchbury J, Barker JN, Trembath RC. Identification of *ZNF313/RNF114* as a novel psoriasis susceptibility gene. Human Molecular Genetics 2008;17(13):1938-1945.
58. Ling H, Waterworth DM, Stirnadel HA, Pollin TI, Barter PJ, Kesäniemi YA, Mahley RW, McPherson R, Waeber G, Bersot TP, Cohen JC, Grundy SM, Mooser VE, Mitchell BD. Genome-wide Linkage

and Association Analyses to Identify Genes Influencing Adiponectin Levels: The GEMS Study. Obesity 2009;17(4):737-744.

59. Raelson JV, Little RD, Ruether A, Fournier H, Paquin B, Eerdewegh PV, Bradley WEC, Croteau P, Nguyen-Huu Q, Segal J, Debrus S, Allard R, Rosenstiel P, Franke A, Jacobs G, Nikolaus S, Vidal J-M, Szego P, Laplante N, Clark HF, Paulussen RJ, Hooper JW, Keith TP, Belouchi A, Schreiber S. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. Proceedings of The National Academy of Sciences of the United States of America 2007;104(37):14747-14752.
60. Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, Bingham SA, Brage S, Smith GD, Ekelund U, Gillson CJ, Glaser B, Golding J, Hardy R, Khaw K-T, Kuh D, Luben R, Marcus M, McGeehin MA, Ness AR, Northstone K, Ring SM, Rubin C, Sims MA, Song K, Strachan DP, Vollenweider P, Waeber G, Waterworth DM, Wong A, Deloukas P, Barroso I, Mooser V, Loos RJ, Wareham NJ. Genetic variation in *LIN28B*is associated with the timing of puberty. Nature Genetics 2009;41(6):729-733.
61. Zhang X-J, Huang W, Yang S, Sun L-D, Zhang F-Y, Zhu Q-X, Zhang F-R, Zhang C, Du W-H, Pu X-M, Li H, Xiao F-L, Wang Z-X, Cui Y, Hao F, Zheng J, Yang X-Q, Cheng H, He C-D, Liu X-M, Xu L-M, Zheng H-F, Zhang S-M, Zhang J-Z, Wang H-Y, Cheng Y-L, Ji B-H, Fang Q-Y, Li Y-Z, Zhou F-S, Han J-W, Quan C, Chen B, Liu J-L, Lin D, Fan L, Zhang A-P, Liu S-X, Yang C-J, Wang P-G, Zhou W-M, Lin G-S, Wu W-D, Fan X, Gao M, Yang B-Q, Lu W-S, Zhang Z, Zhu K-J, Shen S-K, Li M, Zhang X-Y, Cao T-T, Ren W, Zhang X, He J, Tang X-F, Lu S, Yang J-Q, Zhang L, Wang D-N, Yuan F, Yin X-Y, Huang H-J, Wang H-F, Lin X-Y, Liu J-J. Psoriasis genome-wide association study identifies susceptibility variants within *LCE* gene cluster at 1q21. Nature Genetics 2009;41(2):205-210.

62. Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W, Xhu X-O. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1 Nature Genetics 2009;41(3):324-328.

63. Nakano M, Ikeda Y, Taniguchi T, Yagi T, Fuwa M, Omi N, Tokuda Y, Tanaka M, Yoshii K, Kageyama M, Naruse S, Matsuda A, Mori K, Kinoshita S, Tashiro K. Three susceptible loci association with primary open-angle glaucoma identified by genome-wide association study in a Japanese population. Proceedings of The National Academy of Sciences of the United States of America 2009;106(31):12838-12842.

64. Sanna S, Busonero F, Maschio A, McArdle PF, Usala G, Dei M, Lai S, Mulas A, Piras MG, Perseu L, Masala M, Marongiu M, Crisponi L, Naitza S, Galanello R, Abecasis GR, Shuldiner AR, Schlessinger D, Cao A, Uda M. Common variants in the *SLCO1B3* locus are association with bilirubin levels and unconjugated hyperbilirubinemia. Human Molecular Genetics 2009;18(14):2711-2718.

65. Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ. A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Colour and Skin Pigmentation. Public Library of Science: Genetics 2008;4(5):e1000074.

66. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tasi Y-Y, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nature Genetics 2008;40(5):616-622.

67. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, Bueno-de-Mesquita HB, Gross M, Helzlsouer K, Jacobs EJ, LaCroix A, Zheng W, Albanes D, Bamlet W, Berg CD, Berrino F, Bingham S, Buring JE, Bracci PM, Canzian F, Clavel-Chapelon F, Clipp S, Cotterchio

M, de Andrade M, Duell EJ, Fox Jr. JW, Gallinger S, Gaziano JM, Giovannucci EL, Goggins M, Gonzàlez CA, Hallmans G, Hankinson SE, Hassan M, Holly EA, Hunter DJ, Hutchinson A, Jackson R, Jacobs KB, Jenab M, Kaaks R, Klein AP, Kooperberg C, Kurtz RC, Li D, Lynch SM, Mandelson M, McWilliams RR, Mendelsohn JB, Michaud DS, Olson SH, Overvad K, Patel AV, Peeters PHM, Rajkovic A, Riboli E, Risch HA, Shu X-O, Thomas G, Tobias GS, Trichopoulos D, Van Den Eeden SK, Virtamo J, Wactawski-Wende J, Wolpin BM, Yu H, Yu K, Zeleniuch-Jacquotte A, Chanock SJ, Hartge P, Hoover RN. Genome-wide association study identifies variants in the *ABO* locus associated with susceptibility to pancreatic cancer. Nature Genetics 2009;41(9):986-990.

68. Bishop DT, Demenais F, Iles MM, Harland M, Taylor JC, Corda E, Randerson-Moor J, Aitken JF, Avril M-F, Azizi E, Bakker B, Bianchi-Scarrà G, Bressac-de Paillerets B, Calista D, Cannon-Albright LA, Chin-A-Woeng T, Dębniak T, Galore-Haskel G, Ghiorzo P, Gut I, Hansson J, Hočevar M, Höiom V, Hopper JL, Ingvar C, Kanetsky PA, Kefford RF, Landi MT, Lang J, Lubiński J, Mackie R, Malvehy J, Mann GJ, Martin NG, Montgomery GW, van Nieuwpoort FA, Novakovic S, Olsson H, Puig S, Weiss M, van Workum W, Zelenika D, Brown KM, Goldstein AM, Gillanders EM, Boland A, Galan P, Elder DE, Gruis NE, Hayward NK, Lathrop GM, Barrett JH, Newton Bishop JA. Genome-wide association study identifies three loci associated with melanoma risk. Nature Genetics 2009;41(8):920-925.

69. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, members of the CORGI Consortium, Kerr D, Gray R, Peto J, Cazier J-B, Tomlinson I, Houlston RS. A genome-wide association study shows that common alleles of *SMAD7* influence colorectal cancer risk. Nature Genetics 2007;39(11):1315-1317.

70. Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, Amos CI, Houlston RS. Deciphering the Impact of Common Genetic Variation on Lung Cancer Risk: A Genome-Wide Association Study. Cancer Research 2009;69(16):6633-6641.

71. Brown KM, MacGregor S, Montgomery GW, Craig DW, Zhao ZZ, Iyadurai K, Henders AK, Homer N, Campbell MJ, Stark M, Thomas S, Schmid H, Holland EA, Gillanders EM, Duffy DL, Maskiell JA, Jetann J, Ferguson M, Stephan DA, Cust AE, Whiteman D, Green A, Olsson H, Puig S, Ghiorzo P, Hansson J, Demenais F, Goldstein AM, Gruis NA, Elder DE, Bishop JN, Kefford RF, Giles GG, Armstrong BK, Aitken JF, Hopper JL, Martin NG, Trent JM, Mann GJ, Hayward NK. Common sequence variants on 20q11.22 confer melanoma susceptibility. Nature Genetics 2008;40(7):838-840.

72. Capasso M, Devoto M, Hou C, Asgharzadeh S, Glessner JT, Attiyeh EF, Mosse YP, Kim C, Diskin SJ, Cole KA, Bosse K, Diamond M, Laudenslager M, Winter C, Bradfield JP, Scott RH, Jagannathan J, Garris M, McConville C, London WB, Seeger RC, Grant SFA, Li H, Rahman N, Rappaport E, Hakonarson H, Maris JM. Common variants in *BARD1* influence susceptibility to high-risk neuroblastoma. Nature Genetics 2009;41 (6):718-723.

73. Cui R, Kamatani Y, Takahashi A, Usami M, Hosono N, Kawaguchi T, Tsunoda T, Kamatani N, Kubo M, Nakamura Y, Matsuda K. Functional Variants in *ADH1B* and *ALDH2* Coupled With Alcohol and Smoking Synergistically Enhance Esophageal Cancer Risk. Gastroenterology 2009;137(5):1768-1775.

74. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, Sullivan K, Vijayakrishnan J, Wang Y, Pittman AM, Sunter NJ, Hall AG, Dyer MJS, Matutes E, Dearden C, Mainou-Fowler T, Jackson GH, Summerfield G, Harris RJ, Pettitt AR, Hillmen P, Allsup DJ, Bailey JR, Pratt G, Pepper C, Fegan C, Allan JM, Catovsky D, Houlston RS. A genome-wide association

study identifies six susceptibility loci for chronic lymphocytic leukemia. Nature Genetics 2008;40(10):1204-1210.

75. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, Wiklund F, Robbins C, Isaacs SD, Cheng Y, Li G, Sun J, Chang B-L, Marovich L, Wiley KE, Bälter K, Stattin P, Adami H-O, Gielzak M, Yan G, Sauvageot J, Liu W, Kim JW, Bleecker ER, Meyers DA, Trock BJ, Partin AW, Walsh PC, Isaacs WB, Grönberg H, Xu J, Carpten JD. Two Genome-wide Association Studies of Aggressive Prostate Cancer Implicate Putative Prostate Tumor Supressor Gene *DAB2IP*. Journal of the National Cancer Institute 2007;99(24):1836-1844.
76. Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, the SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen C-Y, Wu P-E, Wang H-C, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo K-Y, Noh D-Y, Ahn S-H, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low Y-L, Bogdanova N, Schürmann P, Dörk T, Tollenaar RAEM, Jacobi CE, Devilee P, Klijn JGM, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MWR, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko Y-D, Spurdle AB, Beesley J, Chen X, kConFab, AOCS Management Group, Mannermaa A, Kosma V-M, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BAJ. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447(7148):1087-1093.

77. Eeles RA, Kote-Jarai Z, Giles GG, Al Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, The UK Genetics Prostate Cancer Study Collaborators, British Association of Urological Surgeons' Section of Oncology, The UK ProtecT Study Collaborators, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF. Multiple newly identified loci associated with prostate cancer susceptibility. Nature Genetics 2008;40(3):316-321.

78. Galvan A, Falvella FS, Spinola M, Frullanti E, Leoni VP, Noci S, Alonso MR, Zolin A, Spada E, Milani S, Pastorino U, Incarbone M, Santambrogio L, Neira AG, Dragani TA. A polygenic model with common variants may predict lung adenocarcinoma risk in humans. International Journal of Cancer 123(10):2327-2330.

79. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, Kosarin K, Olsh A, Bergeron J, Ellis NA, Klein RJ, Clark AG, Norton L, Dean M, Boyd J, Offit K. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proceedings of The National Academy of Sciences of the United States of America 2008;105(11):4340-4345.

80. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So W-Y, Tong PCY, Ng MCY,

Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CNA, Rotimi C, Chan JCN, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Two variants on chromosome 17 confer prostate cancer risk, and the one in *TCF2* protects against type 2 diabetes. Nature Genetics 2007;39(8):977-983.

81. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Marcos JG-I, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsson GV, Mayordomo JI, Kiemeney LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nature Genetics 2007;39(5):631-637.
82. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, Kristinsson KT, Birgisdottir B, Ghosh S, Magnusdottir DN, Thorlacius S, Thorleifsson G, Zheng SL, Sun J, Chang B-L, Elmore JB, Breyer JP, McReynolds KM, Bradley KM, Yaspan BL, Wiklund F, Stattin P, Lindström S, Adami H-O, McDonnell SK, Schaid DJ, Cunningham JM, Wang L, Cerhan JR, St. Sauver JL, Isaacs SD, Wiley KE, Partin AW, Walsh PC, Polo S, Ruiz-Echarri M, Navarrete S, Fuertes F, Saez B, Godino J, Weijerman PC, Swinkels DW, Aben KK, Witjes JA, Suarez BK, Helfand BT, Frigge ML, Kristjansson K, Ober C, Jonsson E, Einarsson GV, Xu J, Gronberg H, Smith JR,

Thibodeau SN, Isaacs WB, Catalona WJ, Mayordomo JI, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nature Genetics 2008;40(3):281-283.

83. Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygsdottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, Wahlfors T, Tammela T, Breyer JP, McReynolds KM, Bradley KM, Saez B, Godino J, Navarrete S, Fuertes F, Murillo L, Polo E, Aben KK, van Oort IM, Suarez BK, Helfand BT, Kan D, Zanon C, Frigge ML, Kristjansson K, Gulcher JR, Einarsson GV, Jonsson E, Catalona WJ, Mayordomo JI, Kiemeney LA, Smith JR, Schleutker J, Barkardottir RB, Kong A, Thorsteinsdottir U, Rafnar T, Stefansson K. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nature Genetics 2009;41(10):1122-1126.
84. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, He H, Blondal T, Geller F, Jakobsdottir M, Magnusdottir DN, Matthiasdottir S, Stacey SN, Skarphedinsson OB, Helgadottir H, Li W, Nagy R, Aguillo E, Faure E, Prats E, Saez B, Martinez M, Eyjolfsson GI, Bjornsdottir US, Holm H, Kristjansson K, Frigge ML, Kristvinsson H, Gulcher JR, Jonsson T, Rafnar T, Hjartarsson H, Mayordomo JI, de la Chapelle A, Hrafnkelsson J, Thorsteinsdottir U, Kong A, Stefansson K. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. Nature Genetics 2009;41(4):460-464.
85. COGENT Study. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nature Genetics 2008;40(12):1426-1435.
86. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowka J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, Chen C, Goodman G, Field JK, Liloglou T, Xinarianos G, Cassidy A, McLaughlin J, Liu G, Narod S, Krokan HE, Skorpen F, Elvestad MB, Hveem K, Vatten L, Linseisen J, Clavel-Chapelon F, Vineis P, Bueno-

de-Mesquita HB, Lund E, Martinez C, Bingham S, Rasmuson T, Hainaut P, Riboli E, Ahrens W, Benhamou S, Lagiou P, Trichopoulos D, Holcátová I, Merletti F, Kjaerheim K, Agudo A, Macfarlane G, Talamani R, Simonato L, Lowry R, Conway DI, Znaor A, Healy C, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008;452(7187):633-637.

87. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RC, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni Jr. JF, Hoover RN, Thomas G, Chanock SJ. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. Nature Genetics 2007;39(6):870-874.

88. Kanetsky PA, Mitra N, Vardhanabhuti S, Li M, Vaughn DJ, Letrero R, Ciosek SL, Doody DR, Smith LM, Weaven J, Albano A, Chen C, Starr JR, Rader DJ, Godwin AK, Reilly MP, Hakonarson H, Schwartz SM, Nathanson KL. Common variation in *KITLG* and at 5q31.3 predisposes to testicular germ cell cancer. Nature Genetics 2009;41(7):811-815.

89. Kibriya MG, Jasmine F, Argos M, Andrulis IL, John EM, Chang-Claude J, Ahsan H. A pilot genome-wide association study of early-onset breast cancer. Breast Cancer Research and Treatment 2009;114(3):463-477.

90. Kiemeney LA, Thorlacius S, Sulem P, Geller F, Aben KKH, Stacey SN, Gudmundsson J, Jakobsdottir M, Bergthorsson JT, Sigurdsson A, Blondal T, Witjes JA, Vermeulen SH, Hulsberden-van de Kaa CA, Swinkels DW, Ploeg M, Cornel EB, Vergunst H, Thorgeirsson TE, Gudbjartsson D, Gudjonsson SA, Thorleifsson G, Kristinsson KT, Mouy M, Snorradottir S, Placidi D, Campagna M, Arici C, Koppova K, Gurzau E, Rudnai P, Kellen E, Polidoro S, Guarrera S, Sacerdote C, Sanchez M, Saez B,

Valdivia G, Ryk C, de Verdier P, Lindblom A, Golka K, Bishop DT, Knowles MA, Nikulasson S, Petursdottir V, Jonsson E, Geirsson G, Kristjansson B, Mayordomo JI, Steineck G, Porru S, Buntinx F, Zeegers MP, Fletcher T, Kumar R, Matullo G, Vineis P, Kiltie AE, Gulcher JR, Thorsteinsdottir U, Kong A, Rafnar T, Stefansson K. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nature Genetics 2008;40(11):1307-1312.

91. Liu P, Vikis HG, Wang D, Lu Y, Wang Y, Schwartz AG, Pinney SM, Yang P, de Andrade M, Petersen GM, Wiest JS, Fain PR, Gazdar A, Gaba C, Rothschild H, Mandal D, Coons T, Lee J, Kupert E, Seminara D, Minna J, Bailey-Wilson JE, Wu X, Spitz MR, Eisen T, Houlston RS, Amos CI, Anderson MW, You M. Familial Aggregation of Common Sequence Variants on 15q24-25.1 in Lung Cancer. Journal of the National Cancer Institute 2008;100(18):1326-1330.
92. Maris JM, Mosse YP, Bradfield JP, Hou C, Monni S, Scott RH, Asgharzadeh S, Attiyeh EF, Diskin SJ, Laudenslager M, Winter C, Cole KA, Glessner JT, Kim C, Frackelton EC, Casalunovo T, Eckert AW, Capasso M, Rappaport EF, McConville C, London WB, Seeger RC, Rahman N, Devoto M, Grant SFA, Li H, Hakonarson H. Chromosome 6p22 Locus Associated with Clinically Aggressive Neuroblastoma. The New England Journal of Medicine 2008;358(24):2585-2593.
93. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njølstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubiński J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A, EPIC Study, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Matsuda F, Blanche H, Gut I, Heath S, Lathrop M, Brennan P. Lung cancer susceptibility locus at 5p15.33. Nature Genetics 2008;40(12):1404-1406.

94. Murabito JM, Rosenberg CL, Finger D, Kreger BE, Levy D, Splansky GL, Antman K, Hwang S-J. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. BMC Medical Genetics 2007;8(Suppl 1):S6.

95. Murabito JM, Rosenberg CL, Finger D, Kreger BE, Levy D, Splansky GL, Antman K, Hwang S-J. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. BMC Medical Genetics 2007;8(Suppl 1):S6.

96. Ng CC, Yew PY, Puah SM, Krishnan G, Yap LF, Teo SH, Lim PVH, Govindaraju S, Ratnavelu K, Sam CK, Takahashi A, Kubo M, Kamatani N, Nakamura Y, Mushiroda T. A genome-wide association study identifies *ITGA9* conferring risk of nasopharyngeal carcinoma. Journal of Human Genetics 2009;54(7):392-397.

97. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JAE, Allan JM, Tomlinson IP, Taylor M, Greaves M, Houlston RS. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nature Genetics 2009;41(9):1006-1010.

98. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, Simon M, Marie Y, Boisselier B, Delattre J-Y, Hoang-Xuan K, El Hallani S, Idbaih A, Zelenika D, Andersson U, Henriksson R, Bergenheim AT, Feychting M, Lönn S, Ahlbom A, Schramm J, Linnebank M, Hemminki K, Kumar R, Hepworth SJ, Price A, Armstrong G, Liu Y, Gu X, Yu R, Lau C, Schoemaker M, Muir K, Swerdlow A, Lathrop M, Bondy M, Houlston RS. Genome-wide association study identifies five susceptibility loci for glioma. Nature Genetics 2009;41(8):899-904.

99. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, Iyadurai K, Becker N, Brooks-Wilson A, Curry JD, Spinelli JJ, Holly EA, Riby J, Zhang L, Nieters A, Smith MT, Brown KM. Genetics variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nature Genetics 2009;41(8):873-875.

100. Song H, Ramus RJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, Anton-Culver H, Chang-Claude J, Cramer DW, DiCioccio R, Dörk T, Goode EL, Goodman MT, Schildkraut JM, Sellers T, Baglietto L, Beckmann MW, Beesley J, Blaakaer J, Carney ME, Chanock S, Chen Z, Cunningham JM, Dicks E, Doherty JA, Dürst M, Ekici AB, Fenstermacher D, Fridley BL, Giles G, Gore ME, De Vivo I, Hillemanns P, Hogdall C, Hogdall E, Iversen ES, Jacobs IJ, Jakubowska A, Li D, Lissowska J, Lubiński J, Lurie G, McGuire V, McLaughlin J, Mędrek K, Moorman PG, Moysich K, Narod S, Phelan C, Pye C, Risch H, Runnebaum IB, Severi G, Southey M, Stram DO, Thiel FC, Terry KL, Tsai Y-Y, Tworoger SS, Van Den Berg DJ, Vierkant RA, Wang-Gohrke S, Webb PM, Wilkens LR, Wu AH, Yang H, Brewster W, Ziogas A, Australian Cancer (Ovarian) Study, The Australian Ovarian Cancer Study Group, The Ovarian Cancer Association Consortium, Houlston R, Tomlinson I, Whittemore AS, Rossing MA, Ponder BAJ, Pearce CL, Ness RB, Menon U, Kjaer SK, Gronwald J, Garcia-Closas M, Fasching PA, Easton DF, Chenevix-Trench G, Berchuck A, Pharoah PDP, Gayther SA. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2 Nature Genetics 2009;41(9):996-1000.

101. Spain SL, Cazier J-B, The CORGI Consortium, Houlston R, Carvajal-Carmona L, Tomlinson I. Colorectal Cancer Risk Is Not Associated with Increased Levels of Homozygosity in a Population from the United Kingdom. Cancer Research 2009;69(18):7422-7429.

102. Spinola M, Leoni VP, Galvan A, Korsching E, Conti B, Pastorino U, Ravagnani F, Columbano A, Skaug V, Haugen A, Dragani TA. Genome-wide single nucleotide polymorphism analysis of lung cancer risk detects the KLF6 gene. Cancer Letters 2007;251(2):311-316.

103. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchard L, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K,

Bergthorsson JT, Kostic J, Frigge ML, Geller F, Gudbjartsson D, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johansson J, Sveinsson T, Myrdal G, Grimsson HN, Jonsson T, von Holst S, Werelius B, Margolin S, Lindblom A, Mayordomo JI, Haiman CA, Kiemeney LA, Johansson OT, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nature Genetics 2007;39(7):865-869.

104. Stacey SN, Gudbjartsson DF, Sulem P, Bergthorsson JT, Kumar R, Thorleifsson G, Sigurdsson A, Jakobsdottir M, Sigurgeirsson B, Benediktsdottir KR, Thorisdottir K, Ragnarsson R, Scherer D, Rudnai P, Gurzau E, Koppova K, Höiom V, Botella-Estrada R, Soriano V, Juberías P, Grasa M, Carapeto FJ, Tabuenca P, Gilaberte Y, Gudmundsson J, Thorlacius S, Helgason A, Thorlacius T, Jonasdottir A, Blondal T, Gudjonsson SA, Jonsson GF, Saemundsdottir J, Kristjansson K, Bjornsdottir G, Sveinsdottir SG, Mouy M, Geller F, Nagore E, Mayordomo JI, Hansson J, Rafnar T, Kong A, Olafsson JH, Thorsteinsdottir U, Stefansson K. Common variants on 1q36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. Nature Genetics 2008;40(11):1313-1318.
105. Sun J, Zheng SL, Wiklund F, Isaacs SD, Li G, Wiley KE, Kim S-T, Zhu Y, Zhang Z, Hsu F-C, Turner AR, Stattin P, Liu W, Kim JW, Duggan D, Carpten J, Isaacs W, Grönberg H, Xu J, Chang B-L. Sequence Variants at 22q13 Are Associated with Prostate Cancer Risk. Cancer Research 2009;69(1):10-15.
106. Tenesa A, Farrington SM, Prendergast JGD, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJL, Smith LA, Kavoussanakis K, Koessler T, Pharoah PDP, Buch S, Schafmayer C, Tepel J, Schreiber S, Völzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IPM, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda

K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CMT, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H, Dunlop MG. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nature Genetics 2008;40(5):631-637.

107. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cussenot O, Valeri A, Andriole GL, Crawford ED, Tucker M, Gerhard DS, Fraumeni Jr. JF, Hoover R, Hayes RB, Hunter DJ, Chanock SJ. Multiple loci identified in a genome-wide association study of prostate cancer. Nature Genetics 2008;40(3):310-315.

108. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee I-M, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni Jr. JF, Hoover RN, Chanock SJ, Hunter DJ. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). Nature Genetics 2009;41(5):579-584.

109. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, the CORGI Consortium, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier J-B, Houlston R. A genome-wide

association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nature Genetics 2007;39(8):984-988.

110. Tomlinson IPM, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijayakrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JGD, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, Gray R, Maher ER, Lucassen A, Kerr D, Evans DGR, The CORGI Consortium, Schafmayer C, Buch S, Völzke H, Hampe J, Schreiber S, John U, Koessler T, Pharoah P, van Wezel T, Morreau H, Wijnen JT, Hopper JL, Southey MC, Giles GG, Severi G, Castellví-Bel S, Ruiz-Ponte C, Carracedo A, Castells A, The EPICOLON Consortium, Försti A, Hemminki K, Vodicka P, Naccarati A, Lipton L, Ho JWC, Cheng KK, Sham PC, Luk J, Agúndez JAG, Ladero JM, de la Hoya M, Caldés T, Nittymäki I, Tuupanen S, Karhu A, Aaltonen L, Cazier J-B, Campbell H, Dunlop MG, Houlston RS. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nature Genetics 2008;40(5):623-630.

111. Treviño LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, Willman C, Neale G, Downing J, Raimondi SC, Pui C-H, Evans WE, Relling MV. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nature Genetics 2009;41(9):1001-1005.

112. Tse K-P, Su W-H, Chang K-P, Tsang N-M, Yu C-J, Tang P, See L-C, Hsueh C, Yang M-L, Hao S-P, Li H-Y, Wang M-H, Liao L-P, Chen L-C, Lin S-R, Jorgensen TJ, Chang Y-S, Shugart YY. Genome-wide Association Study Reveals Multiple Nasopharyngeal Carcinoma-Associated Loci within the HLA Region at Chromosome 6p21.3. The American Journal of Human Genetics 2009;85(2):194-203.

113. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Sptiz MR, Eisen T, Amos CI, Houlston RS. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nature Genetics 2008;40(12):1407-1409.

114. Wrensch M, Jenkins RB, Chang JS, Yeh R-F, Xiao Y, Decker PA, Ballman KV, Berger M, Buckner JC, Chang S, Giannini C, Halder C, Kollmeyer TM, Kosel ML, LaChance DH, McCoy L, O'Neill BP, Patoka J, Pico AR, Prados M, Quesenberry C, Rice T, Rynearson AL, Smirnov I, Tihan T, Wiemels J, Yang P, Wiencke JK. Variants in the *CDKN2B* and *RTEL1* regions are associated with high-grade glioma susceptibility. Nature Genetics 2009;41(8):905-908.

115. Wu X, Ye Y, Kiemeney LA, Sulem P, Rafnar T, Matullo G, Seminara D, Yoshida T, Saeki N, Andrew AS, Dinney CP, Czerniak B, Zhang Z-f, Kiltie AE, Bishop DT, Vineis P, Porru S, Buntinx F, Kellen E, Zeegers MP, Kumar R, Rudnai P, Gurzau E, Koppova K, Mayordomo JI, Sanchez M, Saez B, Lindblom A, de Verdier P, Steineck G, Mills GB, Schned A, Guarrera S, Polidoro S, Chang S-C, Lin J, Chang DW, Hale KS, Majewski T, Grossman HB, Thorlacius S, Thorsteinsdottir U, Aben KKH, Witjes JA, Stefansson K, Amos CI, Karagas MR, Gu J. Genetic variation in the prostate stem cell antigen gene *PSCA* confers susceptibility to urinary bladder cancer. Nature Genetics 2009;41(9):991-995.

116. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni Jr. JF, Hoover R, Hunter DJ, Chanock SJ, Thomas G. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nature Genetics 2007;39(5):645-649.

117. Zanke BW, Greenwood CMT, Rangrej J, Kufsta R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier J-F, Robidoux F, Sladek R, Montpetit A, Campbell P, Bezieau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous MEM, Campbell H, Blanche H, Sahbatou M, Tubacher E, Bonaiti-Pellié C, Buecher B, Riboli E, Kury S, Chanock SJ,

Potter J, Thomas G, Gallinger S, Hudson TJ, Dunlop MG. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nature Genetics 2007;39(8):989-994.

118. Zheng W, Long J, Gao Y-T, Li C, Zheng Y, Xiang Y-B, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, Lu W, Shu X-O. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nature Genetics 2009;41(3):324-328.
119. Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Dononvan MC, Williams J, Owen MJ, Kirov G. A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Medical Genomics 2008;1(44).
120. Anney RJL, Lasky-Su J, Ó'Dúshláine C, Kenny E, Neale BM, Mulligan A, Franke B, Zhou K, Chen W, Christiansen H, Arias-Vásquez A, Banaschewski T, Buitelaar J, Ebstein R, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen H, Asherson P, Faraone SV, Gill M. Conduct Disorder and ADHD: Evaluation of Conduct Problems as a Categorical and Quantitative Trait in the International Multicentre ADHD Genetics Study. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1369-1378.
121. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Jamra RA, Höfels S, Propping P, Satagopan J, NIMH Genetics Initiative Bipolar Disorder Consortium, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Molecular Psychiatry 2008;13(2):197-207.

122. Beecham GW, Martin ER, Li Y-J, Slifer MA, Gilbert JR, Haines JL, Pericak-Vance MA. Genome-wide Association Study Implicates a Chromosome 12 Risk Locus for Late-Onset Alzheimer Disease. The American Journal of Human Genetics 2009;84(1):35-43.

123. Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, Schjeide BMM, Hooli B, DiVito J, Ionita I, Jiang H, Laird N, Moscarillo T, Ohlsen KL, Elliott K, Wang X, Hu-Lince D, Ryder M, Murphy A, Wagner SL, Blacker D, Becker KD, Tanzi RE. Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to *APOE*. The American Journal of Human Genetics 2008;83(5):623-632.

124. Blauw HM, Velding JH, van Es MA, van Vught PW, Saris CGJ, van der Zwaag B, Franke L, Hurbach PH, Wokke JH, Ophoff RA, van den Berg LH. Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. Lancet Neurology 2008;7(4):319-326.

125. Carrasquillo MM, Zou F, Pankratz VS, Wilcox SL, Ma L, Walker LP, Younkin SG, Younkin CS, Younkin LH, Bisceglio GD, Ertekin-Taner N, Crook JE, Dickson DW, Petersen RC, Graff-Radford NR, Younkin SG. Genetic variation in *PCDH11X* is associated with susceptibility to late-onset Alzheimer's disease. Nature Genetics 2009;41(2):192-198.

126. Chiò A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai S-L, Mora G, Fung H-C, Britton A, Arepalli S, Gibbs JR, Nalls M, Berger S, Kwee LC, Oddone EZ, Ding J, Crews C, Rafferty I, Washecka N, Hernandez D, Ferrucci L, Bandinelli S, Guralnik J, Macciardi F, Torri F, Lupoli S, Chanock SJ, Thomas G, Hunter DJ, Gieger C, Wichmann HE, Calvo A, Mutani R, Battistini S, Giannini F, Caponnetto C, Mancardi GL, La Bella V, Valentino F, Monsurrò MR, Tedeschi G, Marinou K, Sabatelli M, Conte A, Mandrioli J, Sola P, Salvi F, Bartolomei I, Siciliano G, Carlesi C, Orrell RW, Talbot K, Simmons Z, Connor J, Pioro EP, Dunkley T, Stephan DA, Kasperaviciute D, Fisher EM, Jabonka S, Sendtner M, Beck M, Bruijn L, Rothstein J, Schmidt S, Singleton A, Hardy J,

Traynor BJ. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Human Molecular Genetics 2009;18(8):1524-1532.

127. Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L, Halperin RF, Marlowe L, Kaleem M, Walker DG, Ravid R, Heward CB, Rogers J, Papassotiropoulos A, Reiman EM, Hardy J, Stephan DA. A High-Density Whole-Genome Association Study Reveals That *APOE* Is the Major Susceptibility Gene for Sporadic Late-Onset Alzheimer's Disease. 2007;68(4):613-618.
128. Cronin S, Berger S, Ding J, Schymick JC, Washecka N, Hernandez DG, Greenway MJ, Bradley DG, Traynor BJ, Hardiman O. A genome-wide association study of sporadic ALS in a homogenous Irish population. Human Molecular Genetics 2008;17(5):768-774.
129. Cronin S, Tomik B, Bradley DG, Slowik A, Hardiman O. Screening for replication of genome-wide SNP associations in sporadic ALS. European Journal of Human Genetics 2009;17(2):213-218.
130. Dunckley T, Huentelman MJ, Craig DW, Pearson JV, Szelinger S, Joshipura K, Halperin RF, Stamper C, Jensen KR, Letizia D, Hesterlee SE, Pestronk A, Levine T, Bertorini T, Graves MC, Mozaffar T, Jackson CE, Bosch P, McVey A, Dick A, Barohn R, Lomen-Hoerth C, Rosenfeld J, O'Connor DT, Zhang K, Crook R, Ryberg H, Hutton M, Katz J, Simpson EP, Mitsumoto H, Bowser R, Miller RG, Appel SH, Stephan DA. Whole-Genome Analysis of Sporadic Amyotrophic Lateral Sclerosis. The New England Journal of Medicine 2007;357(8):775-788.
131. Ferreira MAR, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre DJ, MacLean AW, St. Clair D, Robinson M, Van Beck M, Pereira ACP, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A,

Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Wellcome Trust Case Control Consortium, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N. Collaborative genome-wide association analysis supports a role for *ANK3* and *CACNA1C* in bipolar disorder. Nature Genetics 2008;40(9):1056-1058.

132. Feulner TM, Laws SM, Friedrich P, Wagenpfeil S, Wurst SHR, Riehle C, Kuhn KA, Krawczak M, Schreiber S, Nikolaus S, Förstl H, Kurz A, Riemenschneider M. Examination of the current top candidate genes for AD in a genome-wide association study. Molecular Psychiatry 2010;15(7):756-766.
133. Fung H-C, Scholz S, Matarin M, Simón-Sánchez J, Hernandez D, Britton A, Gibbs JR, Langefeld C, Stiegert ML, Schymick J, Okun MS, Mandel RJ, Fernandez HH, Foote KD, Rodríguez RL, Peckham E, De Vrieze FW, Gwinn-Hardy K, Hardy JA, Singleton A. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurology 2006;5(11):911-916.
134. Hallmayer J, Faraco J, Lin L, Hesselson S, Winkelmann J, Kawashima M, Mayer G, Plazzi G, Nevsimalova S, Bourgin P, Hong S-C, Honda Y, Honda M, Högl B, Longstreth Jr. WT, Montplaisir J, Kemlink D, Einen M, Chen J, Musone SL, Akana M, Miyagawa T, Duan J, Desautels A, Erhardt C, Hesla PE, Poli F, Frauscher B, Jeong J-H, Lee S-P, Ton TGN, Kvale M, Kolesar L, Dobrovolnà M, Nepom GT, Salomon D, Wichmann H-E, Rouleau GA, Gieger C, Levinson DF, Gejman PV, Meitinger T, Young T, Peppard P, Tokunaga K, Kwok P-Y, Risch N, Mignot E. Narcolepsy is strongly associated with the T-cell receptor alpha locus. Nature Genetics 2009;41(6):708-711.
135. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Love S, Kehoe PG,

Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Schürmann B, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Hüll M, Rujescu D, Goate AM, Kauwe JSK, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel K-H, Klopp N, Wichmann H-E, Carrasquillo MM, Pankratz VS, Younkin SG, Holmans PA, O'Donovan M, Owen MJ, Williams J. Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease. Nature Genetics 2009;41(10):1088-1093.

136. Hattori E, Toyota T, Ishitsuka Y, Iwayama Y, Yamada K, Ujike H, Morita Y, Kodama M, Nakata K, Minabe Y, Nakamura K, Iwata Y, Takei N, Mori N, Naitoh H, Yamanouchi Y, Iwata N, Ozaki N, Kato T, Nishikawa T, Kashiwa A, Suzuki M, Shioe K, Shinohara M, Hirano M, Nanko S, Akahane A, Ueno M, Kaneko N, Watanabe Y, Someya T, Hashimoto K, Iyo M, Itokawa M, Arai M, Nankai M, Inada T, Yoshida S, Kunugi H, Nakamura M, Iijima Y, Okazaki Y, Higuchi T, Yoshikawa T. Preliminary Genome-Wide Association Study of Bipolar Disorder in the Japanese Population. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2009;150B(8):1110-1117.

137. Heinzen EL, Need AC, Hayden KM, Chiba-Falek O, Roses AD, Strittmatter WJ, Burke JR, Hulette CM, Welsh-Bohmer KA, Goldstein DB. Genome-Wide Scan of Copy Number Variation in Late-Onset Alzheimer's Disease. Journal of Alzheimer's Disease 2010;19(1):69-77.

138. Huentelmann MJ, Papassotiropoulos A, Craig DW, Hoerndli FJ, Pearson JV, Huynh K-D, Corneveaux J, Hänggi J, Mondadori CRA, Buchmann A, Reiman EM, Henke K, de Quervain DJF, Stephan DA. Calmodulin-binding transcription activator 1 (*CAMTA1*) alleles predispose human episodic memory performance. Human Molecular Genetics 2007;16(12):1469-1477.

139. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460(7256):748-752.

140. Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC. A genome-wide association study in 574 schizophrenia trios using DNA pooling. Molecular Psychiatry 2009;14(8):796-803.

141. Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fiévet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O, the European Alzheimer's Disease Initiative Investigators, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossù P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanché H, Dartigues J-F, Tzourio C, Gut I, Van Broeckhoven C, Alpérovitch A, Lathrop M, Amouyel P. Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. Nature Genetics 2009;41(10):1094-1099.

142. Landers JE, Melki J, Meininger V, Glass JD, van den Berg LH, van Es MA, Sapp PC, van Vught PWJ, McKenna-Yasek DM, Blauw HM, Cho T-J, Polak M, Shi L, Wills A-M, Broom WJ, Ticozzi N, Silani V, Ozoguz A, Rodriguez-Leyva I, Veldink JH, Ivinson AJ, Saris CGJ, Hosler BA, Barnes-Nessa A, Couture N, Wokke JHJ, Kwiatkowski Jr. TJ, Ophoff RA, Cronin S, Hardiman O, Diekstra FP, Leigh PN, Shaw CE, Simpson CL, Hansen VK, Powell JF, Corcia P, Salachas F, Heath S, Galan P, Georges F, Horvitz HR, Lathrop M, Purcell S, Al-Chalabi A, Brown Jr. RH. Reduced expression of the *Kinesin-Associated Protein 3* (*KIFAP3*) gene increases survival in sporadic amyotrophic lateral sclerosis. Proceedings of The National Academy of Sciences of the United States of America 2009;106(22):9004-9009.

143. Lasku-Su J, Neale BM, Franke B, Anney RJL, Zhou K, Maller JB, Vasquez AA, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen HC, Taylor E, Daly M, Laird N, Lange C, Faraone SV. Genome-Wide Association Scan of Quantitative Traits for Attention Deficit Hyperactivity Disorder Identifies Novel Associations and Confirms Candidate Gene Associations. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1345-1354.

144. Lasky-Su J, Anney RJL, Neale BM, Franke B, Zhou K, Maller JB, Vasquez AA, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberg A, Sergeant J, Sonuga-Barke E, Steinhausen HC, Taylor E, Daly M, Laird N, Lange C, Faraone SV. Genome-Wide Association Scan of the Time to Onset of Attention Defecit Hyperactivity Disorder. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1355-1358.

145. Latourelle JC, Pankratz N, Dumitriu A, Wilk JB, Goldwurm S, Pezzoli G, Mariani CB, DeStefano AL, Halter C, Gusella JF, Nichols WC, Myers RH, Foroud T, PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories. Genomewide association study for onset age in Parkinson disease. BMC Medical Genetics 2009;10(98).

146. Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherlapati R, Malhotra AK. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. Molecular Psychiatry 2007;12(6):572-580.

147. Lesch K-P, Timmesfeld N, Renner TJ, Halperin R, Röser C, Nguyen TT, Craig DW, Romanos J, Heine M, Meyer J, Freitag C, Warnke A, Romanos M, Schäfer H, Walitza S, Reif A, Stephan DA, Jacob C. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. Journal of Neural Transmission 2008;115(11):1573-1585.

148. Li H, Wetten S, Li L, St. Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M, Coletta N, Delisle R, Dhalla D, Ehm MG, Feldman HH, Fornazzari L, Gauthier S, Goodgame N, Guzman D, Hammond S, Hollingworth P, Hsiung G-Y, Johnson J, Kelly DD, Keren R, Kertesz A, King KS, Lovestone S, Loy-English I, Matthews PM, Owen MJ, Plumpton M, Pryse-Phillips W, Prinjha RK, Richardson JC, Saunders A, Slater AJ, St. George-Hyslop PH, Stinnett SW, Swartz JE, Taylor RL, Wherrett J, Williams J, Yarnall DP, Gibson RA, Irizarry MC, Middleton LT, Roses AD. Candidate Single-Nucleotide Polymorphisms From a Genomewide Association Study of Alzheimer Disease. Archives of Neurology 2008;65(1):45-53.

149. Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN, Hoffman JD, Slifer SH, Hedges DJ, Cukier HN, Griswold AJ, McCauley JL, Beecham GW, Wright HH, Abramson RK, Martin ER, Hussman JP, Gilbert JR, Cuccaro ML, Haines JL, Pericak-Vance MA. A Genome-wide Association Study of Autism Reveals a Common Novel Risk Locus at 5p14.1. Annals of Human Genetics 2009;73(Pt 3):263-273.

150. Maranganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, Pant PVK, Frazer KA, Cox DR, Ballinger DG. High-Resolution Whole-Genome Association Study of Parkinson Disease. American Journal of Human Genetics 2005;77(5):685-693.

151. Mead S, Poulter M, Uphill J, Beck J, Whitfield J, Webb TEF, Campbell T, Adamson G, Deriziotis P, Tabrizi SJ, Hummerich H, Verzilli C, Alpers MP, Whittaker JC, Collinge J. Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. Lancet Neurology 2009;8(1):57-66.

152. Melquist S, Craig DW, Huentelman MJ, Crook R, Pearson JV, Baker M, Zismann VL, Gass J, Adamson J, Szelinger S, Corneveaux J, Cannon A, Coon KD, Lincoln S, Adler C, Tuite P, Calne DB, Bigio EH, Uitti RJ, Wszolek ZK, Golbe LI, Caselli RJ, Graff-Radford N, Litvan I, Farrer MJ, Dickson DW, Hutton M, Stephan DA. Identification of a Novel Risk Locus for Progressive Supranuclear

Palsy by a Pooled Genomewide Scan of 500,288 Single-Nucleotide Polymorphisms. The American Journal of Human Genetics 2007;80(4):769-778.

153. Mick E, Neale B, Middleton FA, McGough JJ, Faraone SV. Genome-Wide Association Study of Response to Methylphenidate in 187 Children With Attention-Deficit/Hyperactivity Disorder. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1412-1418.

154. Miyagawa T, Kawashima M, Nishida N, Ohashi J, Kimura R, Fujimoto A, Shimada M, Morishita S, Shigeta T, Lin L, Hong S-C, Faraco J, Shin Y-K, Jeong J-H, Okazaki Y, Tsuji S, Honda M, Honda Y, Mignot E, Tokunaga K. Variant between *CPT1B* and *CHKB* associated with susceptibility to narcolepsy. Nature Genetics 2008;40(11):1324-1328.

155. Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ, Antoniades A, Domenici E, Perry J, Rothen S, Vandeleur CL, Mooser V, Waeber G, Vollenweider P, Preisig M, Lucae S, Müller-Myhsok B, Holsboer F, Middleton LT, Roses AD. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. Molecular Psychiatry 2010;15(6):589-601.

156. Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB, Vasquez AA, Asherson P, Chen W, Banaschewski T, Buitelaar J, Ebstein R, Gill M, Miranda A, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Sonuga-Barke E, Mulas F, Taylor E, Laird N, Lange C, Daly M, Faraone SV. Genome-Wide Association Scan of Attention Deficit Hyperactivity Disorder. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1337-1344.

157. Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, Shianna KV, Yoon W, Kasperavičiūtė D, Gennarelli M, Strittmatter WJ, Bonvicini C, Rossi G, Jayathilake K, Cola PA, McEvoy JP, Keefe RSE, Fisher EMC, St. Jean PL, Giegling I, Hartmann AM, Möller H-J, Ruppert A, Fraser G, Crombie C, Middleton LT, St. Clair D, Roses AD, Muglia P, Francks C, Rujescu D, Meltzer HY, Goldstein DB. A

Genome-Wide Investigation of SNPs and CNVs in Schizophrenia. Public Library of Science: Genetics 2009;5(2):e1000373.

158. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CCA, Howie B, Leung H-T, Hartmann AM, Möller H-J, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman PV, Molecular Genetics of Schizophrenia Collaboration, Cichon S, Nöthen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nature Genetics 2008;40(9):1053-1055.
159. Otowa T, Yoshida E, Sugaya N, Yasuda S, Nishimura Y, Inoue K, Tochigi M, Umekage T, Miyagawa T, Nishida N, Tokunaga K, Tanii H, Sasaki T, Kaiya H, Okazaki Y. Genome-wide association study of panic disorder in the Japanese population. Journal of Human Genetics 2009;54(2):122-126.
160. Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, Doheny KF, Gusella JF, Nichols WC, Foroud T, Myers RH, The PSG-PROGENI and GenePD Investigators, Coordinators and Molecular Genetic Laboratories. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. Human Genetics 2009;124(6);593-605.
161. Poduslo SE, Huang R, Huang J, Smith S. Genome Screen of Late-Onset Alzheimer's Extended Pedigrees Identifies TRPC4AP by Haplotype Analysis. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2009;150B(1)50-55.
162. Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, Saykin AJ, Orro A, Lupoli S, Salvi E, Weiner M, Macciardi F for the Alzheimer's Disease Neuroimaging Initiative. Hippocampal Atrophy as a Quantitative Trait in a Genome-Wide Association Study Identifying Novel Suscepibility Genes for Alzheimer's Disease. Public Library of Science: ONE 2009;4(8):e6501.

163. Reiman EM, Webster JA, Myers AJ, Hardy J, Dunckley T, Zismann VL, Joshipura KD, Pearson JV, Hu-Lince D, Huentelman MJ, Craig DW, Coon KD, Liang WS, Herbert RH, Beach T, Rohrer KC, Zhao AS, Leung D, Bryden L, Marlowe L, Kaleem M, Mastroeni D, Grover A, Heward CB, Ravid R, Rogers J, Hutton ML, Melquist S, Petersen RC, Alexander GE, Caselli RJ, Kukull W, Papassotiropoulos A, Stephan DA. *GAB2* Alleles Modify Alzheimer's Risk in *APOE* ε4 Carriers. Neuron 2007;54(5):713-720.

164. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, Kawaguchi T, Tsunoda T, Watanabe M, Takeda A, Tomiyama H, Nakashima K, Hasegawa K, Obata F, Yoshikawa T, Kawakami H, Sakoda S, Yamamoto M, Hattori N, Murata M, Nakamura Y, Toda T. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nature Genetics 2009;41(12):1303-1307.

165. Schormair B, Kemlink D, Roeske D, Eckstein G, Xiong L, Lichtner P, Ripke S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Högl B, Frauscher B, Gschliesser V, Poewe W, Peglau I, Vodicka P, Vávrová J, Sonka K, Nevsimalova S, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann H-E, Holsboer F, Müller-Myhsok B, Meitinger T, Winkelmann J. *PTPRD* (protein tyrosine phosphatase receptor type delta) is associated with restless legs syndrome. Nature Genetics 2008;40(8):946-948.

166. Schymick JC, Scholz SW, Fung H-C, Britton A, Arepalli S, Gibbs R, Lombardo F, Matarin M, Kasperaviciute D, Hernandez DG, Crews C, Bruijn L, Rothstein J, Mora G, Restagno G, Chiò A, Singleton A, Hardy J, Traynor BJ. Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. Lancet Neurology 2007;6(4)322-328.

167. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, Tozzi F, Li JZ, Burmeister M, Absher D, Thompson RC, Francks C, Meng F, Antoniades A, Southwick AM, Schatzberg AF,

Bunney WE, Barchas JD, Jones EG, Day R, Matthews K, McGuffin P, Strauss JS, Kennedy JL, Middleton L, Roses AD, Watson SJ, Vincent JB, Myers RM, Farmer AE, Akil H, Burns DK, Boehnke M. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proceedings of The National Academy of Sciences of the United States of America 2009;106(18):7501-7506.

168. Sha Q, Zhang Z, Schymick JC, Traynor BJ, Zhang S. Genome-wide association reveals three SNPs associated with sporadic amyotrophic lateral sclerosis through a two-locus analysis. BMC Medical Genetics 2009;10(86).

169. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, Dudbridge F, Holmans PA, Whittemore AS, Mowry BJ, Olincy A, Amin F, Cloninger CR, Silverman JM, Buccola NG, Byerley WF, Black DW, Crowe RR, Oksenberg JR, Mirel DB, Kendler KS, Freedman R, Gejman PV. Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 2009;460(7256):753-757.

170. Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hettema JM, An SS, Neale MC, van den Oord EJCG, Kendler KS, Chen X, Boomsma DI, Middeldorp CM, Hottenga JJ, Slagboom PE, Flint J. A whole genome association study of neuroticism using DNA pooling. Moleculary Psychiatry 2008;13(3):302-312.

171. Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Welsh D, Weinberger DR, Sun C, Flint J, Darvasi A. Genome-Wide Association Identifies a Common Variants in the Reelin Gene That Increases the Risk of Schizophrenia Only in Women. Public Library of Science: Genetics 2008;4(2):e28.

172. Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM, Garriock HA, Yokoyama JS, McGrath PJ, Peters EJ, Scheftner WA, Coryell W, Lawson WB, Jancic D, Gejman PV, Sanders AR, Holmans P, Slager SL, Levinson DF, Hamilton SP. Novel loci for major depression identified by

genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. Molecular Psychiatry 2009 Dec 29 [Epub ahead of print].

173. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, Paisan-Ruiz C, Lichtner P, Scholz SW, Hernandez DG, Krüger R, Federoff M, Klein C, Goate A, Perlmutter J, Bonin M, Nalls MA, Illig T, Gieger C, Houldon H, Steffens M, Okun MS, Racette BA, Cookson MR, Foote KD, Fernandez HH, Traynor BJ, Schreiber S, Arepalli S, Zonozi R, Gwinn K, van der Burg M, Lopez G, Chanock SJ, Schatzkin A, Park Y, Hollenbeck A, Gao J, Huang X, Wood NW, Lorenz D, Deuschl G, Chen G, Riess O, Hardy JA, Singleton AB, Gasser T. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nature Genetics 2009;41(12):1308-1312.
174. Sklar P, Smoller JW, Fan J, Ferreira MAR, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PIW, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. Molecular Psychiatry 2008;13(6):558-569.
175. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zöllner S, Schork NJ, Kelsoe JR. Genome-wide association study of bipolar disorder in European American and African American individuals. Molecular Psychiatry 2009;14(8):755-763.
176. Sonuga-Barke EJS, Lasku-Su J, Neale BM, Oades R, Chen W, Franke B, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Anney R, Miranda A, Mulas F, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Thompson M, Asherson P, Faraone SV. Does Parental Expressed Emotion

Moderate Genetic Effects in ADHD? An Exploration Using a Genome Wide Association Scan. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1359-1368.

177. Sonuga-Barke EJS, Lasky-Su J, Neale BM, Oades R, Chen W, Franke B, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Anney R, Miranda A, Mulas F, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Thompson M, Asherson P, Faraone SV. Does Parental Expressed Emotion Moderate Genetic Effects in ADHD? An Exploration Using a Genome Wide Association Scan. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 2008;147B(8):1359-1368.

178. Stefansson H, Rye DB, Hicks A, Petursson H, Ingason A, Thorgeirsson TE, Palsson S, Sigmundsson T, Sigurdsson AP, Eiriksdottir I, Soebech E, Bliwise D, Beck JM, Rosen A, Waddy S, Trotti LM, Iranzo A, Thambisetty M, Hardarson GA, Kristjansson K, Gudmundsson LJ, Thorsteinsdottir U, Kong A, Gulcher JR, Gudbjartsson D, Stefansson K. A Genetic Risk Factor for Periodic Limb Movements in Sleep. The New England Journal of Medicine 2007;357(7):639-647.

179. Stefansson H, Steinberg S, Petursson H, Gustafsson O, Gudjonsdottir IH, Jonsdottir GA, Palsson ST, Jonsson T, Saemundsdottir J, Bjornsdottir G, Böttcher Y, Thorlacius T, Haubenberger D, Zimprich A, Auff E, Hotzy C, Testa CM, Miyatake LA, Rosen AR, Kristleifsson K, Rye D, Asmus F, Schöls L, Dichgans M, Jakobsson F, Benedikz J, Thorsteinsdottir U, Gulcher J, Kong A, Stefansson K .Variants in the sequence of the *LINGO1* gene confers risk of essential tremor. Nature Genetics 2009;41(3):277-279.

180. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OPH, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O, Nyegaard M, Tuulio-Henriksson A, Ingason A, Hansen T, Suvisaari J, Lonnqvist J, Paunio T, Børglum AD, Hartmann A, Fink-Jensen A, Nordentoft M, Hougaard D, Norgaard-Pedersen B, Böttcher Y, Olesen J, Breuer R, Möller H-J,

Giegling I, Rasmussen HB, Timm S, Mattheisen M, Bitter I, Réthelyi JM, Magnusdottir BB, Sigmundsson T, Olason P, Masson G, Gulcher JR, Haraldsson M, Fossdal R, Thorgeirsson TE, Thorsteinsdottir U, Ruggeri M, Tosato S, Franke B, Strengman E, Kiemeney LA, GROUP, Melle I, Djurovic S, Abramova L, Kaleda V, Sanjuan J, de Frutos R, Bramon E, Vassos E, Fraser G, Ettinger U, Picchioni M, Walker N, Toulopoulou T, Need AC, Ge D, Yoon JL, Shianna KV, Freimer NB, Cantor RM, Murray R, Kong A, Golimbet V, Carracedo A, Arango C, Costas J, Jönsson EG, Terenius L, Agartz I, Petursson H, Nöthen MM, Rietschel M, Matthews PM, Muglia P, Peltonen L, St. Clair D, Goldstein DB, Stefansson K, Collier DA. Common variants conferring risk for schizophrenia. Nature 2009;460(7256):744-747.

181. Sullivan PF, Lin D, Tzeng J-Y, van den Oord E, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman J, Close SL. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Molecular Psychiatry 2008;13(6):570-584.

182. Sullivan PF, de Geus EJC, Willemsen G, James MR, Smit JH, Zandbelt T, Arolt V, Baune BT, Blackwood D, Cichon S, Coventry WL, Domschke K, Farmer A, Fava M, Gordon SD, He Q, Heath AC, Heutink P, Holsboer F, Hoogendijk WJ, Hottenga JJ, Hu Y, Kohli M, Lin D, Lucae S, MacIntyre DJ, Maier W, McGhee KA, McGuffin P, Montgomery GW, Muir WJ, Nolen WA, Nöthen MM, Perlis RH, Pirlo K, Posthuma D, Rietschel M, Rizzu P, Schosser A, Smit AB, Smoller JW, Tzeng J-Y, van Dyck R, Verhage M, Zitman FG, Martin NG, Wray NR, Boomsma DI, Penninx BWJH. Genome-wide association for major depressive disorder : a possible role for the presynaptic protein piccolo. Molecular Psychiatry 2009;14(4):359-375.

183. van den Oord EJCG, Kuo P-H, Hartmann AM, Webb T, Möller H-J, Hettema JM, Giegling I, Bukszár J, Rujescu D. Genomewide Association Analysis Followed by a Replication component Implicates a Novel Candidate Gene for Neuroticism. Archives of General Psychiatry 2008;65(9):1062-1071.

184. van Es MA, Van Vught PW, Blauw HM, Franke L, Saris CG, Andersen PM, Van Den Bosch L, de Jong SW, van 't Slot R, Birve A, Lemmens R, de Jong V, Baas F, Schelhaas HJ, Sleegers K, Van Broeckhoven C, Wokke JHJ, Wijmenga C, Robberecht W, Velding JH, Ophoff RA, van den Berg LH. ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome wide association study. Lancet Neurology 2007;6(10):869-877.

185. van Es MA, van Vught PW, Blauw HM, Franke L, Saris CGJ, Van Den Bosch L, de Jong SW, de Jong V, Baas F, van't Slot R, Lemmens R, Schelhaas HJ, Birve A, Sleegers K, Van Broeckhoven C, Schymick JC, Traynor BJ, Wokke JHJ, Wijmenga C, Robberecht W, Andersen PM, Veldink JH, Ophoff RA, van den Berg LH. Genetic variation in *DPP6* is associated with susceptibility to amyotrophic lateral sclerosis. Nature Genetics 2008;40(1):29-31.

186. van Es MA, Veldink JH, Saris CGJ, Blauw HM, van Vught PWJ, Birve A, Lemmens R, Schelhaas HJ, Groen EJN, Huisman MHB, van der Kooi AJ, de Visser M, Dahlberg C, Estrada K, Rivadeneira F, Hofman A, Zwarts MJ, van Doormaal PTC, Rujescu D, Strengman E, Giegling I, Muglia P, Tomik B, Slowik A, Uitterlinden AG, Hendrich C, Waibel S, Meyer T, Ludolph AC, Glass JD, Purcell S, Cichon S, Nöthen MM, Wichmann H-E, Schreiber S, Vermeulen SHHM, Kiemeney LA, Wokke JHJ, Cronin S, McLaughlin RL, Hardiman O, Fumoto K, Pasterkamp RJ, Meininger V, Melki J, Leigh PN, Shaw CE, Landers JE, Al-Chalabi A, Brown Jr. RH, Robberecht W, Andersen PM, Ophoff RA, van den Berg LH. Genome-wide association study identifies 19p13.3 (*UNC13A*) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nature Genetics 2009;41(10):1083-1087.

187. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, Stray SM, Rippey CF, Roccanova P, Makarov V, Lakshmi B, Findling RL, Sikich L, Stromberg T, Merriman B, Gogtay N, Butler P, Eckstrand K, Noory L, Gochman P, Long R, Chen Z, Davis S, Baker C, Eichler EE, Meltzer PS, Nelson SF, Singleton AB, Lee MK,

Rapoport JL, King M-C, Sebat J. Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. Science 2008;320(5875):539-543.

188. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PMA, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Canton RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger Jr. JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SFA, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009;459(7246):528-533.
189. Weiss LA, Arking DE, The Gene Discovery Project of Johns Hopkins & the Autism Consortium. A genome-wide linkage and association scan reveals novel loci for autism. Nature 2009;461(7265):802-808.
190. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447(7145):661-678.
191. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, Fulda S, Pütz B, Eckstein G, Hauk S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Peglau I, Eisensehr I, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann H-E, Holsboer F, Müller-Myhsok B, Meitinger T. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nature Genetics 2007;39(8):1000-1006.
192. Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T, Nievergelt C, Barrett TB, McKinney R, Schork N, Smith EN, Bloss C, Nurnberger J, Edenberg HJ, Foroud T, Sheftner W, Lawson WB, Nwulia EA, Hipolito M, Coryell W, Rice J, Byerley W, McMahon F, Schulze TG,

Berrettini W, Potash JB, Belmonte PL, Zandi PP, McInnis MG, Zöllner S, Craig D, Szelinger S, Koller D, Christian SL, Liu C, Gershon ES. Singleton deletions throughout the genome increase risk of bipolar disorder. Molecular Psychiatry 2009;14(4):376-380.

193. Peters DL, Barber RC, Flood EM, Garner HR, O'Keefe GE. Methodologic quality and genotyping reproducibility in studies of tumor necrosis factor -308 G-->A single nucleotide polymorphism and bacterial sepsis: implications for studies of complex traits. Critical Care Medicine 2003;31(6):1691-1696.
194. Clark MF, Baudouin SV. A systematic review of the quality of genetic association studies in human sepsis. Journal of Intensive Care Medicine 2006;32(11):1706-1712.
195. Bogardus Jr. ST, Concato J, Feinstein AR. Clinical epidemiology quality in molecular genetic research: the need for methodological standards. Journal of the American Medical Association 1999;281(20):1919-1926.
196. Gluud LL. Bias in clinical intervention research. American Journal of Epidemiology 2006;163(6):493-501.
197. Centers for Disease Control and Prevention: National Office of Public Health Genomics. Public Population Project in Genomics: HuGE Literature Update. Available at http://www.cdc.gov/genomics/hugenet/whatsnew/current.htm#evidence/. Accessed November 8, 2008.
198. The Cochrane Collaboration: The Cochrane Collaboration open learning material. Publication Bias. Available at http://www.cochrane-net.org/openlearning/html/mod15-2.htm. Accessed April 12, 2011.
199. Dickersin K. The existence of publication bias and risk factors for its occurrence. Journal of the American Medical Association 1990;263(10):1385-1389.

200. Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. Journal of the American Medical Association 1990;263(10):1392-1395.

201. Stern JM, Stimes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. British Medical Journal 1997;315(7109):640-645.

202. Emerson GB, Warme WJ, Wolf FM, Heckman JD, Brand RA, Leopold SS. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. Archives of Internal Medicine 2010;170(21):1934-1939.

203. Dickersin K, Chalmers I. Recognising, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the World Health Organization. JLL Bulletin: Commmentaries on the history of treatment evaluation. Available at http://www.jameslindlibrary.org/illustrating/articles/recognising-investigating-and-dealing-with-incomplete-and-biase.pdf. Accessed April 18, 2011.

204. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database of Systematic Reviews, Issue 1. Art. No.: MR000006. DOI: 10.1002/14651858.MR000006.pub3.

205. Rochon PA, Gurwitz JH, Simms RW, Fortin PR, Felson DT, Minaker KL, Chalmers TC. A study of manufacturer supported trials of non-steroidal anti-inflammatory drugs in the treatment of arthritis. Archives of Internal Medicine 1994;154(2):157-163.

206. Tramèr MR, Reynolds DJ, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. British Medical Journal 1997;315(7109):635-640.

207. von Elm E, Poglia G, Walder B, Tramèr MR. Different patterns of duplicate publication: an analysis of articles used in systematic reviews. Journal of the American Medical Association 2004;291(8):974-980.

208. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study pulication bias and outcome reporting bias. Public Library of Science: ONE 2008;3(8):e3081.

209. Sterm JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. British Medical Journal 1997;315(7109):640-645.

210. Dickersin K, Olson CM, Rennie D, Cook D, Flanagin A, Zhu Q, Reiling J, Pace B. Association between time interval to publication and statistical significance. Journal of the American Medical Association 2002;287(21):2829-2831.

211. Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. Cochrane Database of Systematic Reviews, Issue 2. Art. No.: MR000011. DOI: 10.1002/14651858.MR000011.pub2.

212. Egger M, Zellweger-ZÄhner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. Lancet 1997;350(9074):326-329.

213. JÜni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. International Journal of Epidemiology 2002;31(1):123-124.

214. GØtzsche PC. Reference bias in reports of drug trials. British Medical Journal (Clinical Research edition) 1987;295(6599):654-656.

215. Ravnskov U. Frequency of citation and outcome of cholesterol lowering trials. British medical Journal 1992;305(6850):717.

216. Ravnskov U. Quotation bias in reviews of the diet-heart idea. Journal of Clinical Epidemiology 1995;48(5):713-719.

217. Kjaergard LL, Gluud C. Citation bias of hepato-biliary randomized clinical trials. Journal of Clinical Epidemiology 2002;55(4):407-410.

218. Schmidt LM, Gøtzsche PC. Of mites and men: reference bias in narrative review articles: a systematic review. The Journal of Family Practice 2005;54(4):334-338.

219. Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. Journal of Clinical Epidemiology 2007;60(9):939-946.

220. National Institutes of Health: National Human Genome Research Institute. Genome-Wide Association Studies: Fact Sheet. Available at http://www.genome.gov/20019523. Accessed April 7, 2011.

221. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of The National Academy of Sciences of the United States of America 2009;106(23):9362-9367.

222. Khoury MJ, Little J, Burke W. Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease. Cary, North Carolina, United States of America: Oxford University Press, USA, 2003.

223. Khoury M, Bedrosian S, Gwinn M, Higgins J, Ioannidis J, Little J. Human Genome Epidemiology, 2nd Edition: Building the evidence for using genetic information to improve health and prevent disease. Cary, North Carolina, United States of America: Oxford University Press, USA, 2010.

224. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL,

Mackay TFC, McCarroll SA, Visscher PM. Finding the missing heritability in complex diseases. Nature 2009;461(7265):747-753.

225. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. American Journal of Human Genetics 2008;82(1):100-112.

226. McClellan J, King MC. Genetic heterogeneity in human disease. Cell 2010;141(2):210-217.

227. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews: Genetics 2010;11(6):415-425.

228. Orozco G, Barrett JC, Zeggini E. Synthetic associations in the context of genome-wise association scan signals. Human Molecular Genetics 2010;19(R2):R137-R144.

229. Lee W, Bindman J, Ford T, Glozier N, Moran P, Stewart R, Hotopf M. Bias in psychiatric case-control studies: literature survey. The British Journal of Psychiatry 2007;190(3):204-209.

230. Yu W, Wulf A, Yesupriya A, Clyne M, Khoury MJ, Gwinn M. HuGE Watch: tracking trends and patterns of published studies of genetic association and human genome epidemiology in near-real time. European Journal of Human Genetics 2008;16(9):1155-1158.

231. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews: Genetics 2008;9(5):356-369.

232. Newton HJ. 1999. sbe28: Meta-analysis of p-values. *Stata Technical Bulletin* 49: 15-17. Available at http://www.stata.com/support/faqs/res/stb.html. Accessed June 13, 2011.

233. Gaut BS, Long AD. The Lowdown on Linkage Disequilibrium. *Plant Cell* 2003;15(7):1502-1506.

234. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at http://www.mrc-

bsu.cam.ac.uk/cochrane/handbook/index.htm#chapter_9/9_2_7_expressing_intervention_effects_on_log_scales.htm.

235. Khoury MJ, Little J, Gwinn M, Ioannidis JPA. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *International Journal of Epidemiology* 2007;36(2):439-445.

236. Khoury MJ, Millikan R, Little J, Gwinn M. The emergence of epidemiology in the genomics age. *International Journal of Epidemiology* 2004;33(5):936-944.

237. Farrall M, Morris AP. Gearing up for genome-wide gene-association studies. *Human Molecular Genetics* 2005;14(suppl 2):R157-R162.

238. Wacholder S, Chanock S, Garcia-Closas M, El ghormli L, Rothman N. Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies. *Journal of the National Cancer Insitute* 2004;96(6):434-442.

239. Lawrence RW, Evans DM, Cardon LR. Prospects and pitfalls in whole genome association studies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 2005;360(1460):1589-1595.

240. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology* 2011;22(4):450-456.

241. Marcus E. Taming supplemental material. *Cell* 2009;139(1):11.

242. Marcus E. 2010: a publishing odyssey. *Cell* 2010;140(1):9.

243. HuGENet: HuGE Navigator (version 2.0). GWAS Integrator. Available at http://hugenavigator.net/HuGENavigator/gWAHit.do?query=&typeSubmitAll=+Statistics+&Mysubmit=simple&geneOrderType=geneA. Accessed August 26, 2011.