## ARTICLE

# Establishment of genetic associations for complex diseases is independent of early study findings

Thomas A Trikalinos[1,2], Evangelia E Ntzani[1], Despina G Contopoulos-Ioannidis[1,3,4] and John PA Ioannidis*[,1,2,5]

[1]*Clinical and Molecular Epidemiology Unit and Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece;* [2]*Department of Medicine, Tufts-New England Medical Center, Institute for Clinical Research and Health Policy Studies, Tufts University School of Medicine, Boston, USA;* [3]*Department of Paediatrics, University of Ioannina School of Medicine, Ioannina, Greece;* [4]*Department of Pediatrics, George Washington University School of Medicine and Health Sciences, Washington, DC, USA;* [5]*Biomedical Research Institute, Foundation for Research and Technology – Hellas, Ioannina, Greece*

**Numerous genetic association studies for complex diseases are performed. Investigators place emphasis on formal statistical significance (*P*-values <0.05), but the predictive ability of early statistically significant ('positive') findings is unclear. We scrutinized 55 cumulative meta-analyses of genetic associations (579 studies), in order to assess whether having statistical significance in the earliest (first) published study or in at least half among several (⩾3) early-published studies, or high statistical significance in early studies had any predictive ability for establishing or refuting the presence of the genetic association in subsequent research. In 35 associations, a first study was 'positive' and in 15 associations more than half of the early-published reports were 'positive'. The average publication rate of subsequent studies increased 1.71-fold with a 'positive' first report. When compared against the summary results of subsequent research, sensitivity and specificity were 0.65 and 0.38 for the first reports, and 0.40 and 0.73, respectively, when at least three early studies were considered. First studies also had poor predictive ability, when we considered the estimated attributable fraction and coverage of the 95% confidence interval thereof or higher levels of statistical significance. We conclude that although 'positive' findings in the very first reports provide strong incentive for conducting more studies on a putative genetic epidemiological association, the statistical significance or even the magnitude of the effect of early studies cannot adequately predict eventual establishment of an association. Conversely, many genuine epidemiological associations would be missed, if research were abandoned after early underpowered 'negative' studies.**

*Correspondence: Dr JPA Ioannidis, Clinical and Molecular Epidemiology Unit and Clinical Trials and Evidence-Based Medicine Unit, Department of Hygiene and Epidemiology, University of Ioannnina School of Medicine, University Campus, Ioannina 45110, Greece. Tel: +30 26510 97807; Fax: +30 26510 97867; E-mail: jioannid@cc.uoi.gr
Received 27 February 2004; revised 21 April 2004; accepted 22 April 2004

## Introduction

Genetic association studies are a key method for deciphering the potential relationship between the millions of candidate genetic risk factors[1] and complex multigenetic diseases.[2] Genetic epidemiology has become one of the most prolific fields in epidemiological research. However, concerns have been voiced that newly proposed genetic epidemiological associations are not consistently replicated by subsequent research.[3,4] The estimates of the first

published statistically significant studies on a genetic association are probably inflated compared to the truth,[3,5] following a 'winner's curse' phenomenon, which has been described for linkage studies.[6] The ability of the early-published studies on a probed genetic association to determine whether this association indeed exists or not requires further investigation. Researchers often place a lot of emphasis on formal statistical significance, and even characterize studies as 'positive' or 'negative' based on whether the P-value is <0.05 or not. However, it is unclear as to whether this is justified at all. Important questions may be posed: Is it more likely that subsequent research will establish the importance of an association, if a first report finds formally statistically significant rather than nonsignificant results? Is the predictive ability improved when several early reports are considered? Finally, would the estimated effect size and coverage of the 95% confidence interval in first studies offer better predictive information? We set out to answer these questions in an empirical evaluation of evidence accumulated for 55 gene–disease associations.

## Materials and methods
### Database
We used a published database of 55 meta-analyses of 579 genetic association studies of various disease outcomes. Details of the search strategy, eligibility criteria and the resulting database along with a list of the eligible meta-analyses have been published previously.[3,7] For every meta-analysis, we identified the 'first' published study; if several studies were published during the earliest year of the meta-analysis and there was no way to discern their order, we characterized all of them as 'first'. Algorithms for selecting primary outcomes, genetic contrasts and first studies have been described elsewhere.[4] For every meta-analysis, as 'early-published' studies we selected the 3 first published studies as well as any others published in the same calendar year as the third study. We regarded as 'time span' of a meta-analysis the time interval between the publication of the earliest and the latest included report, unless the last literature search of the meta-analysis also covered subsequent years. If the authors of each meta-analysis did not state the time of the last literature search, we assumed it was 1 year prior to the publication of the meta-analysis.

### Statistical methods of meta-analyses
The odds ratio (OR) was used as the metric of choice. We tested for between-study heterogeneity for each meta-analysis with the $\chi^2$-based $Q$ statistic, which was considered significant for $P<0.10$.[8] We used random effects methods to combine the data. Random effects methods allow that the OR may vary between different studies, and they are acceptable even in the presence of heterogeneity.[8] Evaluation of these meta-analyses for potential heterogeneity

between small and larger studies (occasionally suggestive of publication bias) has been described in detail in a previous publication.[7] The DerSimonian and Laird random effects methods provide an estimate of the between-study variance that is added to the within-study variance of each study. All OR estimates are considered statistically significant at the $P<0.05$ level.

In order to assess the evolution of the accruing evidence on a putative genetic association over time, we performed cumulative meta-analyses. In cumulative meta-analysis studies are ordered by ascending year of publication, and the summary OR, confidence intervals (CIs) and statistical significance status are re-estimated at the end of each calendar year, as new data accumulate.[10]

### Comparative analyses
We compared cumulative meta-analyses of genetic associations where a first study was 'positive' vs those with only 'negative' first studies. We also assessed if meta-analyses where at least half of the early-published studies were 'positive' differed from meta-analyses where more than half of these studies were 'negative'. In a further sensitivity analysis, we also assessed the predictive ability of early-published studies with very low P-values (defined as at least two studies with $P<0.01$, or at least two 'positive' studies, one of which has $P<0.001$). The main evaluations excluded the first or the early studies from the cumulative meta-analyses, respectively. Separate secondary analyses included these studies as well. 'Positive' and 'negative' correspond to P-values $<0.05$ and $\geqslant 0.05$, respectively. These terms simply refer to the level of statistical significance. They should not be misinterpreted as having any relationship with the quality of a study or its results.

We examined whether the average rate of studies published per year during the time span of each meta-analysis depended on the statistical findings of the first or early-published reports. We also assessed whether this publication rate depended on the magnitude of the OR in the first or early-published reports (random effects summary OR of several studies, when applicable). Furthermore, we evaluated whether the final estimate of the strength of an association (OR at the end of the meta-analysis) depended on the statistical findings of the first or early-published reports (Mann–Whitney U-test).

We calculated the sensitivity and specificity of 'positive' first or early-published reports and of early studies with very low P-values, using the statistical significance of the cumulative meta-analysis as the best available 'gold standard' for the presence or not of an association. Sensitivity and specificity should be independent of the prevalence of statistically significant meta-analyses, while the positive and negative predictive values (PPV and NPV, respectively) are affected. From a Bayesian perspective, PPV and NPV represent the posterior probabilities (posterior 'knowledge') that are derived by applying the positive and

negative likelihood ratios (LR+ and LR−, respectively) to the prevalence of 'positive' associations among those probed (a prior probability). The prevalence of statistically significant meta-analyses in our sample is probably an upward biased estimate of the genuine genetic associations among those probed, since genetic associations with one or few 'negative' study results may never be addressed by meta-analysis. Thus, we also imputed PPV and NPV values assuming a much lower prevalence (10%) of genuine genetic associations among those probed.

Finally, we examined the sensitivity and specificity of having an estimated attributable fraction (AF) $\geqslant 2\%$ (based on the coverage of the 95% CI) in the first study (random effects summary when several first reports were available). AF is the proportion of the disease risk that is explained by a risk factor. It can be calculated from the frequency of the risk factor and associated OR by the formula

$$AF = \frac{frequency(OR - 1)}{1 + frequency(OR - 1)}.$$

We calculated the range for the anticipated AF, using the upper and lower boundary of the 95% CIs of the OR. We estimated the frequency of the pertinent genetic risk factors using random effects weighted estimates of the frequency across the control groups of the included studies. We selected a priori a cutoff value of 2% for the AF, since smaller values are unlikely to represent important genetic risk factors at the population level and would not be realistically important to decipher.

Analyses were performed in SPSS 11.0 (SPSS Inc., Chicago, IL), STATA 8.2 (STATA Corp., College Station, TX), Stat-Xact 3.0 (Cytel Corp., Boston, MA), and Meta-analyst (Joseph Lau, Boston, MA). P-values are two-tailed.

## Results
### Characteristics of meta-analyses
A total of 35 meta-analyses had at least one 'positive' first study, while the remaining 20 had only 'negative' first

reports. Two-thirds of the ORs of the 'positive' group (24/35) suggested at least a doubling of the odds of disease susceptibility conferred by a genetic risk factor, while ORs were much smaller in the second group. The two groups did not differ in the length of their time span or sample size. 'Positive' first reports were followed by the publication of more studies than 'negative' first reports (Table 1).

In 15 meta-analyses at least half of the early-published studies were 'positive', while in 33 more than half were 'negative'. Seven meta-analyses were excluded from the calculations, because no further reports had appeared after their early-published studies. Meta-analyses where the majority of early reports were 'positive' had a longer time span, greater median cumulative sample size and more studies compared to the rest (Table 2).

Nine meta-analyses had early-published studies with very low P-values, while 39 did not. The summary ORs of the early-published studies did not differ between these two groups ($P = 0.23$, for the pertinent comparison). The same was true for the time span of the meta-analyses (median 5 years (interquartile range, $IQR = 3$, 7) vs 5 ($IQR = 3$, 7); $P = 0.76$). Meta-analyses with very low early P-values tended to have greater cumulative sample size (median 4230 ($IQR = 2398$, 7362) vs 2389 ($IQR = 1199$, 4178); $P = 0.11$) and more subsequent studies (median 9 ($IQR = 6$, 13) vs 5 ($IQR = 3$, 9); $P = 0.10$). These differences became statistically significant when the early-published studies were also considered in the comparisons ($P = 0.026$ and 0.025 for the two comparisons, respectively).

### Rate of subsequent publications
The average rate of subsequent publications increased 1.71-fold (95% CI = 1.39, 2.10) when a 'positive' first study existed, and became 1.17-fold greater (95% CI = 1.08, 1.27) per doubling of the OR in the first study. When both the statistical significance of the first investigations and the magnitude of their OR were simultaneously considered in multivariate modeling, the average rate of study publica-

**Table 1** Characteristics of meta-analyses with 'positive' and 'negative' first studies

| | 'Positive' first study in meta-analysis | | |
| | Yes, N = 35 (432 studies) | No, N = 20 (147 studies) | P-value |
|---|---|---|---|
| OR of first studies, median (IQR) | 2.69 (1.79, 3.24) | 1.33 (1.10, 2.10) | 0.004 |
| Time span, median (IQR) | 6 (5, 8) | 5 (4, 8) | 0.42 |
| Sample size per meta-analysis, median (IQR) | | | |
|   Subsequent studies only[a] | 3,734 (1,906, 5,269) | 2,421 (892, 4,612) | 0.18 |
|   All studies | 4,091 (2,090, 6,947) | 3,141 (1,236, 5,272) | 0.18 |
| Number of studies per meta-analysis, median (IQR) | | | |
|   Subsequent studies only[a] | 8 (5, 14) | 5 (2, 8) | 0.014 |
|   All studies | 9 (7, 15) | 6 (4, 9) | 0.005 |

IQR: interquartile range; OR: odds ratio for the gene−disease association. Meta-analyses for which there is more than one first study are classified under the 'positive' first study group, if at least one of these reports is 'positive'. For consistency, all genetic contrasts have been coined in a way so that all the 'first' studies have values $\geqslant 1.00$; the larger the OR, the stronger the genetic association with the disease.
[a]Excluding the first studies.

**Table 2** Characteristics of meta-analyses based on whether the majority of the early-published studies are 'positive' or 'negative'

| | At least half of the early-published studies 'positive' | | |
| | Yes, N = 15 (195 studies) | No, N = 33 (359 studies) | P-value |
| --- | --- | --- | --- |
| Summary OR of the early-published studies, median (IQR) | 1.70 (1.45, 2.71) | 1.43 (1.21, 1.73) | 0.089 |
| Time span, median (IQR) | 6 (4, 8) | 4 (3, 6) | 0.047 |
| Number of patients/alleles per meta-analysis, median (IQR) | | | |
| Subsequent studies only[a] | 3,935 (2,398, 6,366) | 2,049 (1,086, 3,921) | 0.024 |
| All studies | 5,162 (4,097, 8,542) | 3,730 (2,011, 6,442) | 0.0046 |
| Number of studies per meta-analysis, median (IQR) | | | |
| Subsequent studies only[a] | 9 (6, 11) | 5 (3, 9) | 0.061 |
| All studies | 13 (9, 15) | 8 (6, 12) | 0.043 |

IQR: interquartile range; OR: odds ratio for the gene–disease association. For consistency, all genetic contrasts have been coined in a way so that all the 'first' studies have values $\geqslant 1.00$; the larger the OR, the stronger the genetic association with the disease.
[a]Excluding the early-published studies.

tions increased 1.60-fold (95% CI = 1.28, 2.00) for a 'positive' initial report, while the magnitude of the OR of the first reports had a borderline effect (1.08-fold increase (95% CI = 0.98, 1.19) per doubling of the corresponding OR).

The average rate of subsequent publications did not increase when at least half of the early-published studies were 'positive' (rate ratio 1.02 (95% CI = 0.82, 1.26)), and did not depend on the summary OR of the early-published studies (rate ratio 0.93, (95% CI = 0.76, 1.14) per doubling of the summary OR). However, the presence of early-published studies with very low P-values increased the rate of subsequent publications by 1.38-fold (95% CI = 1.08, 1.75).

### Subsequent magnitude of the genetic effect
The summary OR of the subsequent investigations did not differ between meta-analyses with a 'positive' first study vs meta-analyses with 'negative' first reports. The median subsequent summary OR among the meta-analyses with 'positive' first studies was 1.22 (IQR = 1.06, 1.45) vs 1.16 (IQR = 0.88, 1.55) when the first investigations were 'negative' (P = 0.52 for the comparison between the two groups) (Figure 1). Similarly, there was no difference in the magnitude of the summary OR among meta-analyses where the majority of the early-published studies were 'positive' (median subsequent OR 1.26 (IQR = 1.13, 1.45)) vs those where the majority was 'negative' (median subsequent OR 1.19 (IQR = 1.06, 1.55); P = 0.65 for the comparison between the two groups) (Figure 1). The same was true for meta-analyses where early-published studies had very low P-values vs meta-analyses without very low early P-values (median subsequent OR 1.35 (IQR = 1.13, 1.34) vs 1.25 (IQR = 1.06, 1.54); P = 0.82).

### Subsequent statistical significance of genetic associations
Excluding the first studies, the calculated sensitivity and specificity of a 'positive' first report were low (0.65. and 0.38, respectively; Table 3). Given the observed prevalence of meta-analyses with 'positive' final results in our sample (23/55, 42%), the PPV and the NPV of a 'positive' first study would be 0.43 and 0.60, respectively. If only 10% of the probed genetic associations were genuine, the PPV and NPV of a 'positive' first report would be 0.10 and 0.91, respectively. Including the first studies in the meta-analysis calculations yielded similar results (Table 3).

When the majority of the early-published studies were 'positive', the sensitivity and specificity for predicting statistical significance in the meta-analysis of subsequent research were low (0.40 and 0.73, respectively). However, after the incorporation of the early-published studies in the calculations, sensitivity increased modestly to 0.52 and specificity became very good (0.91) (Table 3). With these more favorable estimates, PPV and NPV would be 0.81 and 0.72, respectively, when the prevalence of genuine associations is 42%. PPV and NPV would be 0.39 and 0.94, when this prevalence is only 10%. The presence of early-published studies with very low P-values had a similar sensitivity and specificity as the presence of a majority of early studies with 'positive' results, when all studies were considered in the meta-analysis calculations. The sensitivity of having very low P-values in early-published studies dropped to 0.20, when the early studies were excluded from the meta-analysis calculations, but the 95% CIs overlapped considerably with the 95% CI of the respective sensitivity of having a majority of 'positive' early reports (Table 3).

### Anticipated AF for the disease risk
The first study of 19 meta-analyses claimed an AF of at least 2% (based on the coverage of the 95% CI), whereas the first study in the remaining 36 could not have such certainty. The synthesis of the subsequent studies claimed an AF of at least 2% in 14 meta-analyses, while another nine meta-analyses eventually excluded an AF of 2%.

Formal statistical significance in the cumulative meta-analyses was not strongly related to whether the first
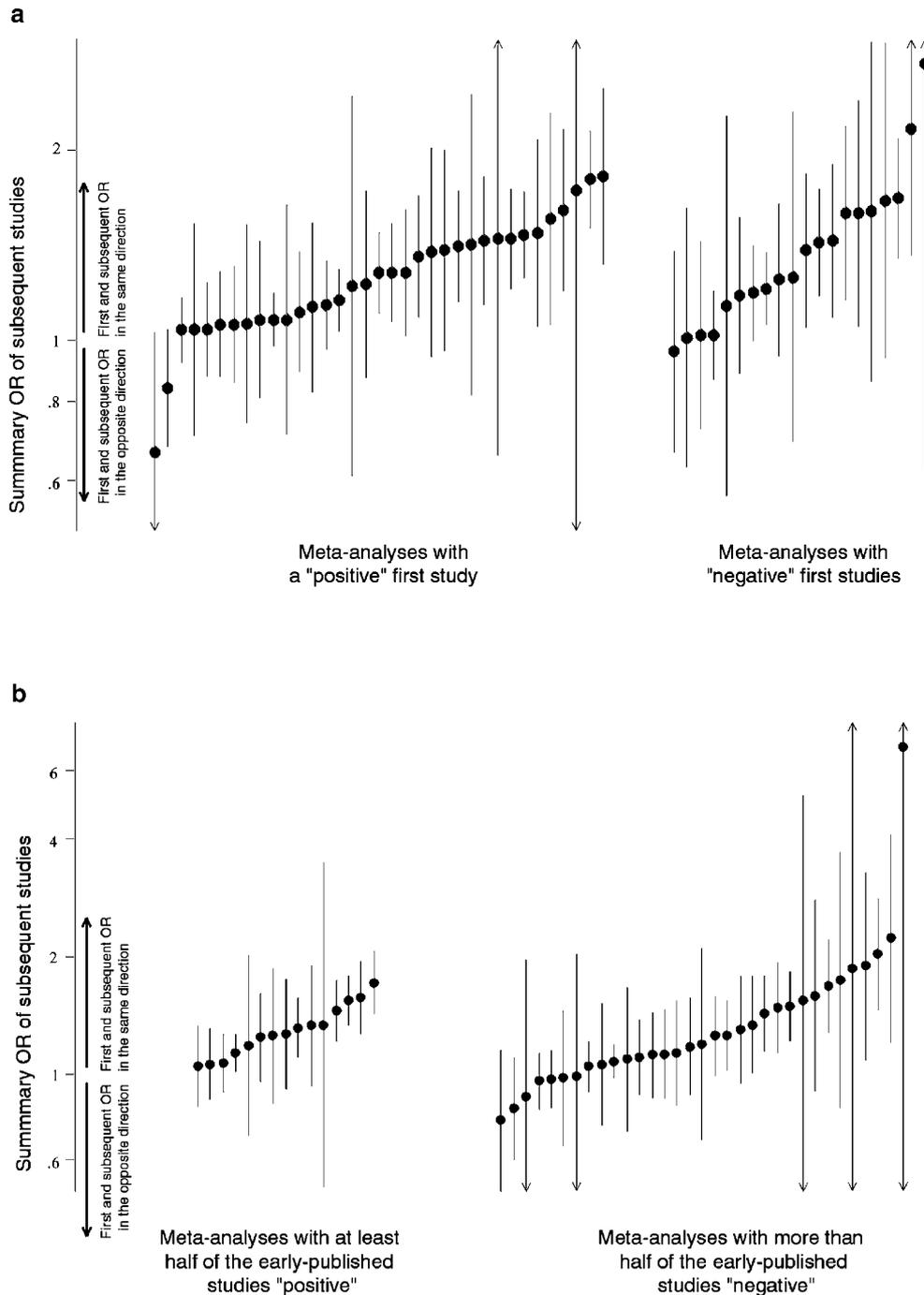
**a**



Meta-analyses with
a "positive" first study

Meta-analyses with
"negative" first studies

**b**



Meta-analyses with at least
half of the early-published
studies "positive"

Meta-analyses with more than
half of the early-published
studies "negative"

**Figure 1** Summary ORs (dots) and corresponding 95% CIs (vertical bars) for research published after the first studies (a) or after at least 3 early-published studies (b) for each one of the eligible meta-analyses (55 for (a) and 48 for (b)). Arrowheads imply that the upper or lower boundary of the 95% CI extend beyond the edges of the graph. Meta-analyses with ORs greater than 1.00 are showing effects in the direction proposed by the first studies, or the synthesis of at least three early-published studies. Meta-analyses with ORs less than 1.00 are showing effects in a direction opposite to that of the first studies, or the synthesis of at least three early-published studies. Ordering is by ascending OR values. Inclusion of the first or early-published studies in the summary OR calculations yielded largely similar results (not shown). OR: odds ratio.

**Table 3** Diagnostic performances of first and early-published studies against the statistical significance of the meta-analysis

| Assessment | Sensitivity (95% CI) | Specificity (95% CI) | LR+ | LR− |
|---|---|---|---|---|
| *First or early-published studies excluded from the meta-analysis* | | | | |
| First study with $P<0.05$ (35/55) | 0.65 (0.43, 0.84) | 0.38 (0.21, 0.56) | 1.1 | 0.92 |
| At least half of early-published studies with $P<0.05$ (15/48) | 0.40 (0.16, 0.68) | 0.73 (0.54, 0.87) | 1.5 | 0.82 |
| Very low $P$-values in early-published studies[a] (9/48) | 0.20 (0.04, 0.48) | 0.82 (0.65, 0.93) | 1.1 | 0.98 |
| Attributable fraction in first study $\geqslant 2\%$ based on 95% CI coverage (19/55) | 0.39 (0.20, 0.61) | 0.69 (0.50, 0.84) | 1.3 | 0.88 |
| *All studies included in the meta-analysis* | | | | |
| First study with $P<0.05$ (35/55) | 0.67 (0.46, 0.83) | 0.39 (0.22, 0.59) | 1.1 | 0.85 |
| At least half of early-published studies with $P<0.05$ (15/48) | 0.52 (0.31, 0.72) | 0.91 (0.72, 0.99) | 5.8 | 0.53 |
| Very low $P$-values in early-published studies[a] (9/48) | 0.47 (0.23, 0.72) | 0.96 (0.78, 1.00) | 11 | 0.55 |
| Attributable fraction in first study $\geqslant 2\%$ based on 95% CI coverage (19/55) | 0.44 (0.25, 0.65) | 0.75 (0.55, 0.89) | 1.8 | 0.75 |

[a]Refers to the presence of $\geqslant 2$ studies with $P$-values $<0.01$, or the presence of $\geqslant 2$ statistically significant studies, one of which has $P<0.001$ among the early-published studies.
CI: confidence interval; LR+: positive likelihood ratio; LR−: negative likelihood ratio.
CIs were derived using exact methods.

studies had claimed or not an AF of at least 2% based on the coverage of the 95% CI. Sensitivity and specificity estimates were modest regardless of whether first and early-published studies were included or not in the calculations (Table 3).

## Discussion

We have shown that the apparent establishment of a proposed genetic epidemiological association is largely independent of the statistical significance found in the first or early-pertinent reports. While 'positive' first reports stimulate the publication of subsequent investigations, they have limited predictive value for proving epidemiological associations on their own. Hence, relying solely on formal significance of first reports – in terms of a $P$-value $<0.05$ – is not an optimal strategy for investing time and resources in order to verify genetic associations. Even when at least three studies are published their results cannot be considered decisive and subsequent research may still give different results. Even early studies with very low $P$-values still have no better predictive ability for the results of subsequent studies. Approaches based on the magnitude of the genetic effect rather than its statistical significance are not necessarily more efficient for detecting associations given the currently employed sample sizes. Even when an early study is not only significant but can also explain 2% of the population disease risk, based on the coverage of the 95% CI, there is no guarantee for the subsequent replication of the findings. Only about a quarter of the meta-analyses could eventually suggest (based on the coverage of the 95% CIs) that they had found a genetic risk factor that would explain at least 2% of the disease risk. Much fewer meta-analyses could exclude a 2% AF with such certainty.

The fact that characteristics of early research cannot clearly identify the associations that are established by future research is not surprising, given that the first published reports often propose inflated and significantly discrepant results compared to subsequent research.[3,5] Given the plethora of genetic information and the theoretically countless possible genetic associations of diseases, an 'auction analogy' has been proposed,[5] where the study with the smaller $P$-values ('most favorable bid') is published first. Nonreplicated early studies could represent incidental spurious findings due to type I error in a setting of multiple comparisons.[11] Under this perspective, the whole phenomenon could be viewed as a meta-analysis-leveled regression to the mean. Biological plausibility may be important to consider when targeting a specific association. However, biological plausibility is usually not straightforward, and it can often be even misleading, especially when evoked *post hoc* to support the epidemiological findings. Conversely, most conducted studies are of a relatively small sample size and, since most associations refer to modest effect sizes (OR 1.20–1.50), they are obviously underpowered.[7,12] Thus type II error is also prominent and important genetic risk factors may be missed if the search is abandoned after limited 'negative' data are obtained.

The average rate of publication of subsequent studies was almost 70% greater, when a first study had reached statistical significance. $P$-values $<0.05$ seem to allure other research teams to probe the gene–disease association proposed by the first study. Formal statistical significance provided a stronger incentive than the actual magnitude of the genetic effect in this regard. We should acknowledge that we examined meta-analyses of published investigations and that some studies may never be published[13,14] or may be published with considerable delay.[15] Thus an alternative explanation may be that not only the conduct of subsequent research but also the ability or intention to publish its results may be compromised when the initial reports on an association are 'negative'. These biases would

pose a threat to the validity of the emerging literature. All well-conducted genetic data should be registered in accessible databases, so that we can obtain a comprehensive picture of the evolving evidence.[16]

In a Bayesian framework one may approach the results of the first or early studies as the data that modify prior probabilities with regard to the presence of a true association into posterior (postdata) probabilities. Even with careful biological targeting of research questions, it is likely that the majority of probed gene–disease pairs do not represent genuine, true associations. Thus the prior probability may be in the 10% range or less. Actually, the extensive adoption of high-throughput techniques with the ability to genotype automatically hundreds and thousands of genetic markers concurrently (the expansion of the 'discovery research' paradigm) means that prior probabilities may become even far smaller than that in the future. The vast majority of probed gene–disease pairs are probably totally unrelated. Our data suggest that the estimated LR+ of early results cannot transform the low prior odds to high enough posterior probabilities. Large-scale validation in several studies with many subjects would be essential, especially if the prior probability is low, as in the setting of 'discovery research' in any field of current molecular medicine.[17,18] Conversely, the estimated LR− does not seem to decrease the likelihood of 'negative' results appreciably and further studies may still be warranted if there is a strong prior biological rationale.

Our analysis was based on meta-analyses retrieved from electronic databases. We have not updated each individual meta-analysis. Updating from our team could have been precarious, since it is often difficult for outside investigators to capture the exact eligibility criteria of the original meta-analysis. Moreover, the statistical significance of a meta-analysis of several studies is not necessarily a 'gold standard' for the eventual establishment of a genetic association. Nevertheless, it is the best available proxy. A more important limitation is that meta-analyses are still conducted for a relatively small, not fully representative, proportion of genetic associations. For some associations only one study is often performed and/or reported, and no subsequent research may be published, especially when findings are 'negative'. Obviously, meta-analyses cannot be conducted in such cases. Hence, 'positive' findings may be over-represented in our database, compared to the plethora of genetic associations that have been or are being assessed. Sensitivity and specificity are nevertheless unbiased in this setting. This also suggests that the effect of a first 'positive' report upon the subsequent rate of publications is probably even stronger than what we estimated. For the large majority of genetic associations, searching is largely a hypothesis-generating exercise and prior beliefs are not so strong to keep pursuing more studies in the face of 'negative' results. Finally, we should acknowledge that in this evaluation we considered only case–control investiga-

tions on unrelated people. Theoretically family-based studies may eliminate the confounding effects of population stratification, but we cannot tell whether they would have better predictive ability for the eventual establishment of an association. Moreover, the large majority of studies in this field use case–control designs.

Our findings could lead to the formulation of some guidelines about the interpretation of early genetic association studies. As in any field of epidemiological research, preliminary, hypothesis-generating studies are useful, but rarely sufficient on their own. The establishment of genetic epidemiological associations requires substantially larger studies, a considerable number of carefully conducted investigations,[11] and all-inclusive meta-analyses.[19] Of course, the interpretation of formal statistical significance in a meta-analysis may entail similar dangers as in single studies. For major diseases further research should be encouraged, if resources can be met, until at least several thousand subjects have been genotyped. Even then there can be limited certainty if the genetic variants are rare. Moreover, the AF may sometimes vary across different settings, for example, racial groups, although most often differences are probably not major.[20] In the presence of major heterogeneity due to population stratification,[21–23] or other factors,[21] continuous monitoring of the updated results on an association may determine both whether further studies are needed, and if so, in what specific population settings. Real-time evidence collection and comprehensive synthesis[24] may be very useful in probing genetic associations of complex diseases, but cautious interpretation is probably always warranted.

### References

1 Zhao Z, Boerwinkle E: Neighboring-nucleotide effects on single nucleotide polymorphisms: a study of 2.6 million polymorphisms across the human genome. *Genome Res* 2002; **12**: 1679–1686.

2 Reich DE, Lander ES: On the allelic spectrum of human disease. *Trends Genet* 2001; **17**: 502–510.

3 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306–309.

4 Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K: A comprehensive review of genetic association studies. *Genet Med* 2002; **4**: 45–61.

5 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. *Nat Genet* 2003; **33**: 177–182.

6 Goring HH, Terwilliger JD, Blangero J: Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 2001; **69**: 1357–1369.

7 Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG: Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003; **361**: 567–571.

8 Lau J, Ioannidis JP, Schmid CH: Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997; **127**: 820–826.

9 DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.

10 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC: Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; **327**: 248–254.

11 Ioannidis JP: Genetic associations: false or true? *Trends Mol Med* 2003; **9**: 135–138.

12 Colhoun HM, McKeigue PM, Davey Smith G: Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; **361**: 865–872.

13 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. *Lancet* 1991; **337**: 867–872.

14 Dickersin K, Min YI: Publication bias: the problem that won't go away. *Ann NY Acad Sci* 1993; **703**: 135–346.

15 Ioannidis JP: Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998; **279**: 281–286.

16 Little J, Bradley L, Bray MS *et al*: Reporting, appraising, and integrating data on genotype prevalence and gene–disease associations. *Am J Epidemiol* 2002; **156**: 300–310.

17 Ransohoff DF: Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer* 2004; **4**: 309–314.

18 Ntzani EE, Ioannidis JPA: Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003; **362**: 1439–1444.

19 Stroup DF, Berlin JA, Morton SC *et al*: Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**: 2008–2012.

20 Wacholder S, Rothman N, Caporaso N: Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 513–520.

21 Vieland VJ: The replication requirement. *Nat Genet* 2001; **29**: 244–245.

22 Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.

23 Gambaro G, Anglani F, D'Angelo A: Association studies of genetic polymorphisms and complex disease. *Lancet* 2000; **355**: 308–311.

24 Khoury MJ, Little J: Human genome epidemiologic reviews: the beginning of something HuGE. *Am J Epidemiol* 2000; **151**: 2–3.