ELSEVIER

Journal of Clinical Epidemiology 58 (2005) 543–549

Journal of Clinical Epidemiology

REVIEW ARTICLES

Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials

John P.A. Ioannidis[a,b,c,*], Thomas A. Trikalinos[a,b]

[a]*Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, MA 02111, USA*

[b]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece*

[c]*Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina 45110, Greece*

Accepted 8 October 2004

Abstract

Background and Objective: Divergent results on the same scientific question generate controversy. We hypothesized that controversial data are attractive to investigators and editors, and thus the most extreme, opposite results would appear very early rather than late, as data accumulate, provided data can be generated rapidly.

Methods: We used data from MEDLINE-indexed meta-analyses of case–control studies on genetic associations (retrospective, hypothesis-generating research with usually rapid turnaround) and meta-analyses of randomized trials of health care interventions (prospective, targeted research that usually takes longer) sampled from the Cochrane Library. Using cumulative meta-analysis, we evaluated how the between-study variance for studies on the same question changed over time and at what point the studies with the most extreme results ever observed had been published.

Results: The maximal between-study variance was more likely to be recorded early in the 44 eligible meta-analyses of genetic associations than in the 37 meta-analyses of health care interventions ($P = .013$). At the time of the first heterogeneity assessment, the most favorable-ever result in support of a specific association was more likely to appear than the least favorable-ever result (22 vs. 10, $P = .017$); the opposite was seen at the second heterogeneity assessment (15 vs. 5, $P = .031$). Such a sequence of extreme opposite results was not seen in the clinical trials meta-analyses. The estimated between-study variance decreased over time in genetic association studies ($P = .010$), but not in clinical trials ($P = .30$).

Conclusion: In contrast to prospective trials, a rapid early sequence of extreme, opposite results is frequent in retrospective hypothesis-generating molecular research.

Keywords: Bias; Meta-analysis; Publication bias; Genetics; Randomized controlled trials

1. Introduction

Diverse investigators may generate data on the same scientific question, and these data may or may not agree among themselves. Although some uncertainty is unavoidable in science, extreme disagreements create confusion and controversy. It would be useful to understand how these controversies arise. An important issue is whether disagreements and the most extreme discrepancies between studies on the same question are more likely to appear early or late, as accumulating data are being published over time.

Scientific research is often subject to publication bias [1,2] and time-lag bias [3]. It is well documented that small studies with nonsignificant results may remain unpublished [1,2], and studies with significant results take a shorter time to be published than studies with nonsignificant results [3]. The more extreme an observed result, the more likely it is to be formally statistically significant and published faster. As a consequence, early publications may present findings that are out of proportion to the truth. If many more studies are performed on the same topic, their results should be, on average, less extreme than the results of the first study [4–6]. Yet is the order of publication of these subsequent studies determined by the results of the first study? We hypothesized that highly contradictory results are most tantalizing and attractive to investigators and editors; thus, they may also have an advantage for rapid publication against

* Corresponding author: Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece. Tel.: +3026510-97807; fax: +3026510-97867.

E-mail address: jioannid@cc.uoi.gr (J.P.A. Iooannidis).

0895-4356/05/$ – see front matter
doi: 10.1016/j.jclinepi.2004.10.019

other results that are closer to the original first publications. In this case, one would expect to see early on in the literature a succession of the most extreme opposite results. Studies with intermediate, potentially less spectacular results may then be published at a slower pace, filling in the gap between the early extremes.

According to our hypothesis, an early rapid succession of extremes would be most common when a large number of studies can be performed worldwide in a short period of time. This may be typical of retrospective study designs, especially in fields with intense research activity. Conversely, this rapid succession of extremes would not be observed for prospective research where studies take considerable time to perform and their conduct is spread over a longer time span. Even if the most contradictory findings have some advantage for more rapid publication, once an extreme result has been reported, this will not suffice to make them appear in the literature before studies with less contradictory findings that have been completed much earlier. To test our hypothesis, we used two large databases of meta-analyses of case–control studies on genetic associations for complex diseases and prospective clinical trials on the efficacy of health care interventions.

2. Methods

2.1. Meta-analysis framework

We used a meta-analytic perspective [7]. A meta-analysis gathers all the studies on the same question, so it is possible to examine when the studies with the most extreme observed results have been published relative to others. In cumulative meta-analysis [8], the results of all studies on a given question are sequentially summarized as they are published over time. Both the summary results and variance estimates may change over time [9]. At the end of each year, summary results are routinely estimated again, whenever there are new pertinent studies that have appeared in the interim. The extent of diversity between the results of the summarized studies (between-study variance) can also be routinely estimated at the same time points. In this framework, one can also examine at which time point the between-study variance is maximized, that is, the maximal diversity exists between the included studies.

Here, we adopted to perform routine assessments of summary effects and between-study variance at the end of each calendar year (when new data were available), to avoid subjective choice on when the accumulating data should be reassessed. Whenever two or more studies had been published in the same calendar year, we did not perform analyses of summary effects and heterogeneity separately after each study, because for studies published in the same calendar year, it is very difficult or even impossible to order them chronologically. The exception is that we separated studies published in the earliest calendar year whenever it was straightforward to identify (as previously described [4]) the "first" study on the specific topic.

2.2. Compared datasets

We investigated two sets of meta-analyses of clinical studies. The full details and data on the meta-analyses are available from the authors upon request. The first dataset pertained to retrospective case–control studies of genetic associations of complex diseases [10,11]. With several million polymorphisms in the human genome [12,13], and with thousands of diseases where multiple gene variants may play a role, an enormous number of potential genetic associations may be postulated and probed [14]. The search for new genetic markers is hypothesis-generating research par excellence. A large number of false positive findings may occur by chance, and many investigators are capable of addressing the same question. Using stored samples, a typical case–control genotyping study may be performed very quickly, even if exceptions do exist (e.g., when questions are strictly predefined and genetic material samples are not readily available and need to be collected).

The second dataset pertained to prospective randomized controlled trials of diverse health care interventions. Controlled clinical trials have been described as experiments that are unwieldy and difficult to conduct [15]. Despite several drawbacks and despite definite exceptions [16], controlled trials usually have *a priori*-defined outcomes, they target specific interventions for specific diseases, and they may take several years from conception to publication [3]. If an extreme result is obtained, subsequent trials may find less spectacular treatment effects, but it is unlikely that the trial with the most contradictory result will be completed and published shortly after the first extreme result. Moreover, it may be considered unethical to conduct further clinical trials, if a large benefit or harm has already been recorded for a specific intervention.

2.3. Genetic associations

Of 55 previously selected cumulative meta-analyses of genetic association studies with binary outcomes indexed in MEDLINE [17], we chose those 44 (534 studies) where (1) at least five studies had been performed, and (2) between-study variance could be assessed in at least three occasions as relevant data accumulated over time ($\geqslant$3 heterogeneity assessments) (Table 1). Eligibility criteria, search strategies, selection of genetic contrasts, and identification of the first studies for the original set of 55 meta-analyses have been described elsewhere [4,17]. For the purposes of this analysis, all genetic contrasts were coined in such a way, so that all the final estimates would consistently have an odds ratio above 1, that is, suggestive of genetic susceptibility.

2.4. Health care interventions

Meta-analyses of controlled clinical trials were obtained from the Cochrane Library [18], a comprehensive database

Table 1
Comparison between meta-analyses of genetic associations and controlled clinical trials

	Meta-analyses of genetic associations	Meta-analyses of controlled clinical trials	*P*-value
Studies, median (IQR)	9 (7–15)	8 (6–14)	.08
Subjects or alleles, median (IQR)	4,082 (2,810–5,912)	1,003 (430–3,343)	<.001
Heterogeneity assessments, median (IQR)	5 (3–6)	5 (4–7)	.22
Years spanned, median (IQR)	6 (5–8)	14 (11–21)	<.001
Statistical significance* at the end of the meta-analysis	23/44	23/37	.50
Statistically significant heterogeneity† at the end of the meta-analysis	23/44	17/37	.66

IQR: Interquartile range, * at the $P < .05$ level, † at the $P < .1$ level.

Meta-analyses of genetic associations are those listed in Table 1 of ref. [17], excluding the ineligible ones (ID 10, 13, 14, 27, 33, 34, 37, 39, 45, 50, 51 in that table). Meta-analyses of controlled clinical trials address pharmacologic interventions ($n = 23$), interventions for smoking cessation ($n = 4$), other cognitive–behavioral treatments ($n = 1$), vaccines ($n = 2$), Chinese herbs ($n = 2$), dietary interventions in pregnancy ($n = 2$), continuous distending pressure ($n = 1$), home-like institutional settings ($n = 1$), and radiotherapy ($n = 1$). A full detailed list is available from the authors upon request.

of systematic reviews on health care interventions. We performed systematic sampling of the 1,519 systematic reviews included in issue 4, 2002, of the Cochrane Library, screening every 10th review in an alphabetical listing. Eligibility criteria were similar to those for the genetics dataset: availability of meta-analysis of binary data comparing an experimental intervention vs. placebo, no treatment or standard treatment with at least five studies and at least three between-study heterogeneity assessments. We selected only the primary outcome for each meta-analysis. When there were several primary outcomes, we selected the one with the maximal number of studies. Outcomes were coined to signify consistently favorable events (e.g., survival, clinical success, etc.). Odds ratios above 1 suggest superiority of the experimental arm). A total of 37 meta-analyses were selected representing 384 clinical trials.

2.5. Analyses

The odds ratio was used as the metric of choice in all meta-analyses. Between-study variance was estimated using the DerSimonian and Laird method [19]. Summary estimates were obtained with the DerSimonian and Laird random effects model for odds ratios. The random effects model allows that results of different studies may differ among themselves [20]. Studies are weighted by the inverse of their variance plus the between-study variance estimate.

For each cumulative meta-analysis we examined whether the maximal between-study variance estimate was found at the first (earliest), second, or subsequent heterogeneity assessments. We similarly recorded at what time in the course of each meta-analysis, each of the two most extreme observed odds ratios ("most favorable-ever" and "least favorable-ever" result) was seen. Finally, we examined whether the between-study variance at the end of the cumulative meta-analysis was larger or smaller than the between-study variance at the time of the first heterogeneity assessment. Data were compared between consecutive heterogeneity assessments in each dataset of meta-analyses, and across the two datasets. Exact tests were used for ordered contingency tables. Paired comparisons for continuous measures used the Wilcoxon matched-pairs signed-rank test.

Analyses were performed in Meta-Analyst (Joseph Lau, Boston, MA), StatXact 3 (Cytel, Boston, MA), and SPSS 11.0 (SPSS, Inc., Chicago, IL). *P*-values are two tailed.

3. Results

3.1. Comparison of meta-analyses

The two sets of meta-analyses had a similar number of studies and heterogeneity assessments, and were not significantly different in the proportion of cases with overall statistically significant odds ratios or between-study heterogeneity. However, as expected, clinical trials had been published over a significantly broader time span than the genetic association studies. Clinical trials also had a smaller sample size (Table 1).

3.2. Early extreme contradictory estimates in genetic associations

In 40 of the 44 meta-analyses of genetic association studies, some between-study variance (estimated variance >0) had been detected at least once during the accumulation of data. In half of these cases (20 of 40), the first (earliest) estimate of between-study variance was the maximal ever observed, while in another 11 cases the second estimate was the maximal ever observed. Thus the most impressive diversity between studies addressing the same research question usually occurred very early during the accumulation of data.

On closer scrutiny it was apparent that early studies often provided most extreme contradictory views for several postulated associations (Fig. 1) [21–23]. In 22 of the 44 meta-analyses, the most extreme estimate ever observed in favor of the genetic association (the most favorable-ever study result) had been already published by the time of the first (earliest) heterogeneity assessment. By the same time, in 10 of the 44 meta-analyses the least favorable-ever study result

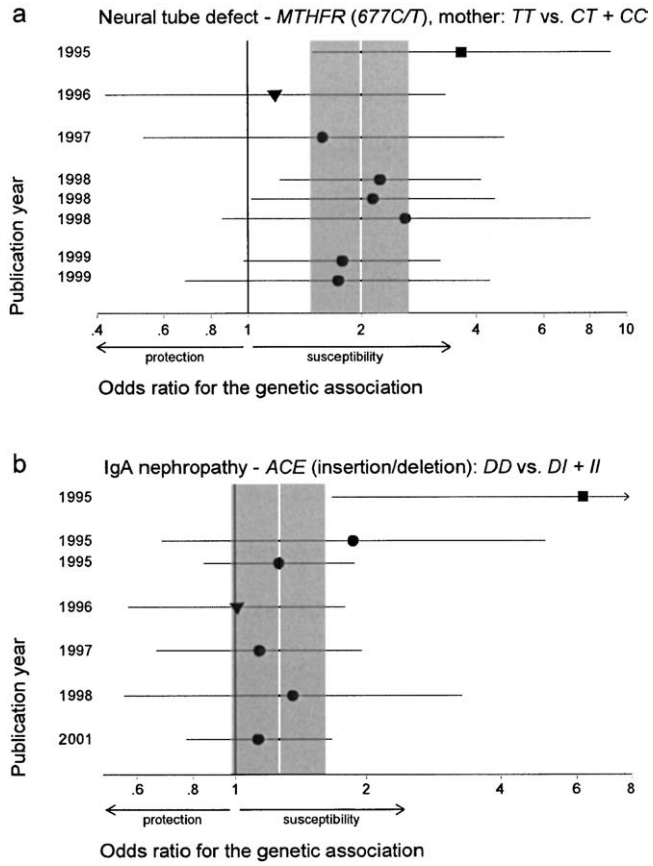

Fig. 1. Extreme differences in the results of a first study vs. a study published shortly thereafter. In both presented examples, studies are ordered chronologically and their results are shown by the odds ratio and 95% confidence intervals. All studies published in the same calendar year are packed together, unless one was clearly the first study. The study with the most favorable-ever results for the presence of an association is shown by a square, and the study with the least favorable-ever results is shown by a triangle, while all other studies are shown by circles. The white line corresponds to the summary odds ratio and the shaded area shows the 95% confidence interval. Also shown is the vertical line of no association (odds ratio = 1). (a) The first published study on the relationship between the methylenetetrahydrofolate reductase (*MTHFR) TT* genotype in the mother and the risk of neural tube defects in the child found a very strong, statistically significant association (odds ratio 3.67, 95% confidence interval 1.47–9.07) and was published in *The Lancet* [21]. The following year, data reported in the same journal [22] showed only a minor nonsignificant trend. Subsequent studies provided intermediate results between these two extremes. (b) The first conducted study on the relationship between the angiotensin converting enzyme *(ACE) DD* genotype and IgA nephropathy showed a highly statistically significant association and was published in the *Journal of Clinical Investigation* [23]. Two other studies published in the same year in nephrology journals found no significant association; a study published the following year found no association at all and between-study variance was maximized. Subsequent studies had intermediate results. The overall data are still inconclusive, but exclude the effect observed in the first study.

had already been published. In five meta-analyses, at the time of the first heterogeneity assessment, both extreme studies (the one with the most favorable-ever results and the one with least favorable-ever results) had already been published. It was more likely for the most favorable-ever result than the least favorable-ever result to be published so early (17 vs. 5 [and 5 ties]; $P = .017$) (Table 2).

At the time of the second heterogeneity assessment, the opposite scenario was seen: the most favorable-ever result appeared in five cases, while the least favorable-ever result appeared in 15 (Table 2). It was significantly more likely for the least favorable-ever result than for the most favorable-ever result to appear at the time of the second heterogeneity assessment (4 vs. 14, $P = .031$; in one case both extremes appeared then).

The deviation between extreme estimates in genetic association studies could be very large as shown in Fig. 2. In 36 of 44 cases (82%), the extremes were even in the completely opposite direction, that is, one extreme study suggested a protective effect while the other extreme suggested increased susceptibility conferred by the same genetic variant.

3.3. Controlled trials of health care interventions

In 32 of 37 meta-analyses some between-study variance (estimated variance >0) had been detected at least once during the accumulation of data. However, the maximal between-study variance did not occur more frequently at the time of the first ($n = 9$) or second ($n = 6$) heterogeneity assessment, rather than in later assessments ($n = 17$).

In controlled trials, the most favorable-ever result and least-favorable-ever result had already been published by the time of the first heterogeneity assessment only in 8 and 14 of the 37 meta-analyses, respectively. In three cases, both extremes had appeared so early. Thus, it was not more likely for the most favorable-ever result than the least favorable-ever result to have been published so early (5 vs. 11 [and 3 ties]; $P = .21$). At the time of the second heterogeneity assessment, the most favorable-ever result appeared in seven cases, while the least favorable-ever result appeared in 5 ($P = .77$) (Table 2).

3.4. Time of occurrence of maximal variance and extreme results in the two types of studies

The maximal between-study variance was significantly more likely to be recorded relatively early in genetic associations than in health care interventions (31 of 40 vs. 15 of 32, respectively, occurred within the first two heterogeneity assessments, $P = .013$).

The most favorable-ever result was more likely to occur very early in genetic association meta-analyses than in meta-analyses of controlled trials ($P = .016$ adjusted for trend). The least favorable-ever result was seen very early equally frequently in genetic association meta-analyses and in meta-analyses of controlled trials ($P = .61$ adjusted for trend) (Table 2).

3.5. Evolution of between-study heterogeneity estimates

In genetic association studies the estimated between-study heterogeneity significantly decreased over time: on average,

Table 2
Time of occurrence of most extreme results in genetic association studies and controlled trials of health care interventions

	Time of the occurrence of most extreme results[a]					
	Most favorable-ever published result			Least favorable-ever published result		
	Early	Middle	Late	Early	Middle	Late
Genetic association studies	22	5	17	10	15	19
Controlled trials of interventions	8	7	22	14	5	18

[a] Early: by the time of the first heterogeneity assessment; Middle: at the time of the second heterogeneity assessment; Late: at any subsequent time.

the latest estimate was smaller than the first estimate of the between-study variance ($P = .010$ by Wilcoxon matched-pairs signed-rank test). This was not the case for health care interventions, where the between-study heterogeneity did not change significantly over time—if anything, the trend was in the opposite direction ($P = .30$) (Fig. 3).

4. Discussion

Publication bias and time lag bias are well-established problems in the biomedical research literature [1–3] and beyond [24–27] and are responsible for the appearance of spuriously favorable results in the peer-reviewed literature. Here we show an additional bias that results in the appearance of extreme, contradictory findings very early during the accumulation of scientific evidence. This phenomenon may occur when a large pool of potential analyses is available and many teams worldwide can generate and disseminate data in limited time. In forefront scientific fields, often many teams may be targeting almost concurrently similar problems. Modern technologies have drastically reduced the time needed to generate experimental data and have inversed the time relationship between data production and data analysis [28]. This progress provides fascinating opportunities for the rapid production and dissemination of scientific information, but may result in an increasing number of studies reporting very contradictory conclusions in short sequence. We suggest that this rapid, early succession of extreme findings may be called the Proteus phenomenon after the mythologic god who rapidly metamorphosed himself to very

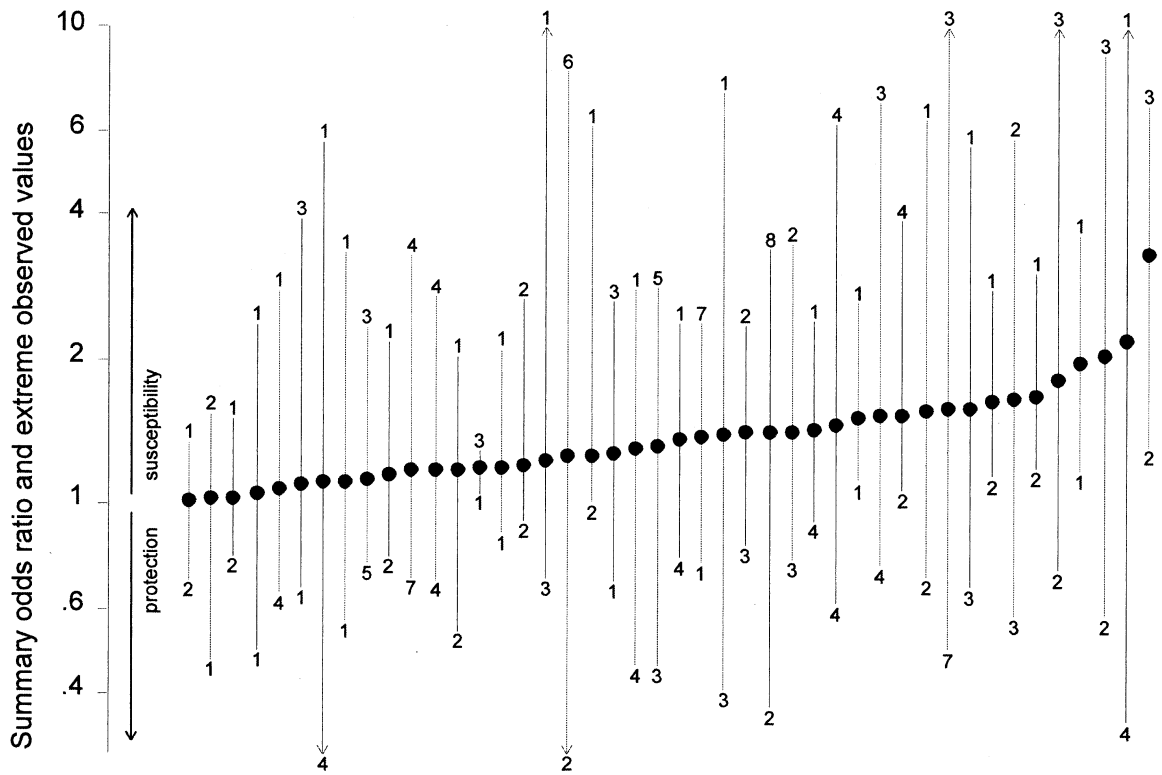

Fig. 2. Range of extreme estimates of the odds ratio observed in various studies on each of 44 postulated genetic associations. Meta-analyses of the 44 associations are ordered with increasing summary odds ratio. For each meta-analysis, the black circles shows the summary odds ratio and the lines extend to the highest (most favorable-ever) and lowest (least favorable-ever) observed odds ratio estimate in any study. The upper number denotes the heterogeneity assessment at which the highest odds ratio estimate was observed. The lower number denotes the heterogeneity assessment at which the lowest odds ratio estimate was observed.

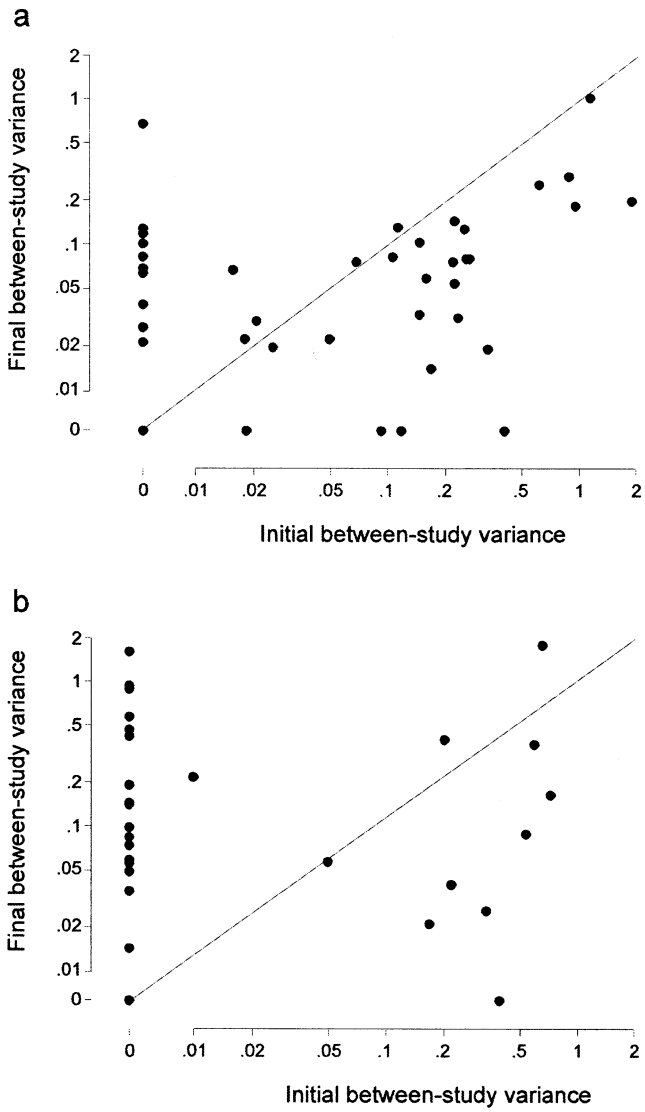

Fig. 3. Estimates of between-study variance at the end of the meta-analysis (considering all studies) are plotted against the respective estimates at the time of the first heterogeneity assessment. The diagonal shows the line where initial and final estimates coincide. Points below the diagonal represent cases where the between-study variance estimate is larger initially (at the time of the first heterogeneity assessment) than finally (including the latest studies). (a) Studies of genetic associations. (b) Trials on health care interventions. One and two outliers are not shown, respectively, in the two panels.

different figures. This bias highlights the importance of performing further validation research and of systematically appraising the evolving accumulated evidence on each research question.

When studies cannot be completed very quickly and research is less exploratory, the Proteus phenomenon may be less prominent. Our observations in a large set of meta-analyses of clinical trials of health care interventions were consistent with this anticipation. Although randomized trials may still be susceptible to publication and time-lag bias [2,3], this bias may be less prominent than for retrospective studies [2]. Extreme discrepancies may appear at any time during the accumulation of randomized evidence, not only in early stages. Actually, corporate pressure and conflicts of interest may also favor the appearance of consistently "positive" results for medical treatments in the early years. It is also conceivable that trials that are conducted late may also be different compared with the earlier ones in terms of the patient population or clinical setting or even their key objectives. When discrepancies occur, they should be carefully scrutinized to understand their clinical meaning [29].

The Proteus phenomenon may, in part, be due to the fact that the pool of genetic association analyses that can be undertaken is enormous [14]. Many of the pursued associations are simply false positive findings due to multiple comparisons. Extreme findings may result by chance, by analytical manipulation, or occasionally by genuine population diversity [5]. Large numbers of gene variants may be probed worldwide, and each may be evaluated for a wide spectrum of clinical disease outcomes and with variable genetic contrasts. Furthermore, genetic associations may sometimes truly differ between various populations [30]. New research findings may draw more attention and get published more rapidly, when extreme effects are detected [11]. However, extreme observations are unlikely to represent the average population-wide truth [6]. If many studies can be performed on the same topic, contradictory results may quickly be generated in another study, with chance, analytical manipulation and/or genuine diversity working in the opposite direction. In high-profile molecular scientific fields such as genetics, highly contradictory results may be attractive for rapid publication. Data may be already available in latent form in large-scale molecular data banks waiting for immediate analysis and publication, once a claim has been made. Meta-analyses would be the appropriate approach for eventually detecting this bias, and it is important that the quality and rigor of meta-analyses in molecular medicine is optimized [31].

We should acknowledge that although we used genetic association studies and randomized trials as prototypes of research designs, no prototype is perfect. For example, some case–control studies may take considerable time to conduct. Conversely, many randomized trials may be conducted concurrently as part of new drug development programs, and the use of surrogate markers has shortened the time it takes to run clinical trials. Furthermore, many clinical trials have very small sample size. Thus, the Proteus phenomenon may become eventually relevant even for prospective, randomized research. However, in the dataset that we used, trials had largely clinical outcomes and were spread over considerable time spans.

We suspect that some of the contradictory results in high throughput scientific domains may actually be already available, when a first study is published. Unfortunately, in our analysis we could not have information on the actual dates of completion of these studies. Further clarification on the

study timing could be obtained only with prospective registries of clinical studies that capture when each study is started and when it is completed [3]. Given the rapid pace of research in several fields in modern molecular medicine, and given the large number of research questions that can be formulated and answered with rapid techniques, we believe that the Proteus phenomenon may become increasingly relevant to the interpretation of research findings.

Acknowledgments

The work was supported by a PENED grant from the General Secretariat for Research and Technology, Greece, and the European Commission. There were no competing financial interests. We thank Evangelia E. Ntzani and Despina G. Contopoulos-Ioannidis for their contribution in background work.

References

[1] Dickersin K, Min YI. Publication bias: the problem that won't go away. Ann N Y Acad Sci 1993;703:135–46.
[2] Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. Lancet 1991;337:867–72.
[3] Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281–6.
[4] Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001; 29:306–9.
[5] Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002;4:45–61.
[6] Siegmund D. Upward bias in estimation of genetic effects. Am J Hum Genet 2002;71:1183–8.
[7] Cooper H, Hedges LV, editors. The handbook of research synthesis. New York: Russell Sage Foundation; 1993.
[8] 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;327:248–54.
[9] Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. Proc Natl Acad Sci USA 2001;98:831–6.
[10] Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.
[11] 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;33:177–82.
[12] 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;12:1679–86.
[13] Sachidanandam R, Weissman D, Schmidt SC, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–33.
[14] Ioannidis JP. Genetic associations: false or true? Trends Mol Med 2003;9:135–8.
[15] Piantadosi S. Clinical trials: A methodological perspective. New York: John Wiley and Sons; 1997.
[16] Jadad AR, Rennie D. The randomized controlled trial gets a middle-aged checkup. JAMA 1998;279:319–20.
[17] Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. Lancet 2003;361:567–71.
[18] The Cochrane Library. Issue 4, available at: http://www.update-software.com. Last accessed on 04/11/2003.
[19] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
[20] Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820–6.
[21] van der Put NM, Steegers-Theunissen RP, Frosst P, et al. Mutated methylenetetrahydrofolate reductase as a risk factor for spina bifida. Lancet 1995;346:1070–1.
[22] Papapetrou C, Lynch SA, Burn J, Edwards YH. Methylenetetrahydrofolate reductase and neural tube defects. Lancet 1996;348:58.
[23] Yoshida H, Mitarai T, Kawamura T, et al. Role of the deletion of polymorphism of the angiotensin converting enzyme gene in the progression and therapeutic responsiveness of IgA nephropathy. J Clin Invest 1995;96:2162–9.
[24] Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull 1979;86:638–41.
[25] Palmer AR. Detecting publication bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. Am Nat 1999;154: 299–332.
[26] Palmer AR. Quasireplication and the contrast of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. Ann Rev Ecol Syst 2000;31:441–80.
[27] Hedges LV, Vevea JL. Estimating effect size under publication bias: small sample properties and robustness of a random effects selection model. J Educ Behav Stat 1996;21:299–332.
[28] Schena M. Microarray analysis. Hoboken, NJ: Wiley-Liss; 2003.
[29] Ioannidis JP, Cappelleri JC, Lau J. Issues in comparisons between meta-analyses and large trials. JAMA 1998;279:1089–93.
[30] Vieland VJ. The replication requirement. Nat Genet 2001;29:244–5.
[31] Attia J, Thakkinstian A, D'Este C. Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. J Clin Epidemiol 2003;56:297–303.