References
1. Sachs JD, McArthur J, Schmidt-Traub G (2005) Response to Amir Attaran. PLoS Med 2: e379. DOI: 10.1371/journal.pmed.0020379
2. Attaran A (2005) An immeasurable crisis? A criticism of the Millennium Development Goals and why they cannot be measured. PLoS Med 2: e318. DOI: 10.1371/journal.pmed.0020318
3. United Nations (2005) Data availability analysis. New York: United Nations. Available: http://unstats.un.org/unsd/mi/techgroup/January2005/Series%20update%20status%20query_FC.xls. Accessed 13 October 2005.
4. Rommelmann V, Setel PW, Hemed Y, Angeles G, Mponezya H, et al. (2005) Cost and results of information systems for health and poverty indicators in the United Republic of Tanzania. Bull World Health Organ 83: 569–577.
5. United Nations Millennium Project (2005) Investing in development: A practical plan to achieve the Millennium Development Goals. New York: Earthscan. 329 p.

Citation: Attaran A (2005) Author's reply. PLoS Med 2(11): e405.

Competing Interests: AA has held small contracts or been paid per diem by the World Bank, the United Nations Development Program, and the Roll Back Malaria Partnership in the last five years. None of these agencies were consulted in the development of this manuscript. Research funding was provided exclusively by the Canada Research Chairs program.

DOI: 10.1371/journal.pmed.0020405

Power, Reliability, and Heterogeneous Results

Ian Shrier

I want to congratulate John P. A. Ioannidis on his thought-provoking Essay [1]. I have two comments.

In Corollary 1, he suggests that small sample sizes mean smaller power, and implies that larger studies with thousands of subjects are more likely to be true. I think it is important to stress that if the effect size is large (e.g., very small variance that is seen in physiological studies), then adequate power is obtained with small numbers. Furthermore, some would argue that exposing subjects to research risks unnecessarily (e.g., when fewer subjects would yield sufficient power) is unethical. Since the analysis is based on power, we should remember that larger is not always better.

In Corollary 4, Ioannidis argues that greater flexibility in designs, definitions, etc. means the results are less likely to be true. I agree that replication of all aspects of the study is more likely to yield consistent results, but this does not necessarily mean true results. Since we don't know a priori which methodological details are most appropriate (e.g., dose, timing, etc.), heterogeneous results from different designs is an important source of information and can lead to a new, more in-depth understanding of the subject—and sometimes even paradigm shifts. I agree with the accompanying Editorial [2] to the article that we need to distinguish between the validity of the data and the validity of the authors' conclusions. ■

Ian Shrier
McGill University
Montreal, Quebec, Canada
E-mail: ian.shrier@mcgill.ca

References
1. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2: e124. DOI: 10.1371/journal.pmed.0020124
2. PLoS Medicine Editors (2005) Minimizing mistakes and embracing uncertainty. PLoS Med 2: e272. DOI: 10.1371/journal.pmed.0020272

Citation: Shrier I (2005) Power, reliability, and heterogeneous results. PLoS Med 2(11): e386.

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020386

The Clinical Interpretation of Research

Stephen G. Pauker

John P. A. Ioannidis emphasizes the central role of prior probabilities [1]. His conclusion rests on the presumed low probability that a hypothesis was true before the study.

Unfortunately, his formulation relates the post-study probability that the study's conclusion is true to the pre-study odds. The results might have been clearer had he also plotted the relation of odds to probability, a curvilinear relationship, assuming the study carried no information. Further, the various graphs are right-truncated at pre-study odds, R, of 1.0 (a probability of 0.5), although his examples go as high as $R = 2.0$. A positive study must, by definition, increase the likelihood that the hypothesis is true. It might have been clearer had Ioannidis chosen to relate odds to odds or probability to probability; in both cases, a neutral study would produce a straight line along a 45-degree diagonal.

The pre-study to post-study relation can more simply be expressed using the odds-likelihood form of Bayes rule—i.e., the post-study odds equal the pre-study odds multiplied times the likelihood ratio (LR) of the study. Then, the equations for positive predictive value (PPV) become the simple product of $R \times \mathrm{LR}$. For a single unbiased study, $\mathrm{LR} = (1 - \beta)/\alpha$. When incorporating study bias, u, as defined by Ioannidis, $\mathrm{LR} = (1 - \beta[1 - u])/(\alpha[1 - u] + u)$. For a typical study with $\alpha = 0.05$ and $\beta = 0.2$ (i.e., with a power of 0.8), LR = 16. When R is less than 1:16 (a probability of 0.0588), the post-study odds will be less than one—i.e., the study's hypothesis will be more likely false than true.

For non-Bayesians, statistical significance testing presumes uninformative prior probability—i.e., $R = 1$. Then, LR would merely need to exceed one for the study's conclusions to be more likely true than false. At the common significance levels (α) of 0.05 and 0.01, the requisite study powers would merely need to exceed 0.05 and 0.01 respectively, corresponding to maximum type II error rates (β) of 0.95 and 0.99. Such lax requirements would almost always be met for a published study. Hence, the common belief that the vast majority of studies have valid conclusions would be correct if we can assume that the pre-study odds are truly uninformative. However, as Ioannidis suggests, this is unlikely to be the case.

Two more corollaries might be added. The higher the pre-study odds that the study's hypothesis is true, the lower the requisite power (study size and effect size) required to make the study's findings more likely true than false. When studies are published, the investigator should estimate the pre-study odds and report the LR implied by the observed effect.

From the perspective of an epidemiologist or a statistician, the relevant question is whether the study's hypothesis is true—i.e., is the probability of the hypothesis greater than 0.5? For clinicians and their patients, the relevant question is whether a particular strategy should be followed in an

individual patient or a subset of similar patients. That decision (or recommendation to the patient) will depend on the pre-study likelihood of benefit in that patient and on the relative magnitude of benefits and risks of that strategy, if the diagnosis in that patient is uncertain. For many such decisions, the "more likely true than false" criterion may not be the best decision rule. For serious diseases and treatments of only modest risk, post-study probabilities of considerably less than 0.5 may be sufficient to justify treatment [2].

Ioannidis's provocative Essay is a timely call for careful consideration of published studies. The odds-likelihood formulation suggested herein may be helpful in providing a more intuitive model. Clinicians now need to take it to the next step. ■

Stephen Pauker
Tufts-New England Medical Center
Boston, Massachusetts, United States of America
E-mail: spauker@tufts-nemc.org

References
1. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2: e124. DOI: 10.1371/journal.pmed.0020124
2. Pauker SG, Kassirer JP (1975) Therapeutic decision making: A cost-benefit analysis. N Eng J Med 293: 229–234.

Citation: Pauker S (2005) The clinical interpretation of research. PLoS Med 2(11): e395.

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020395

Truth, Probability, and Frameworks

Jonathan D. Wren

James T. Kirk: Harry lied to you, Norman. Everything Harry says is a lie. Remember that, Norman: Everything he says is a lie.
Harry Mudd: Now I want you to listen to me very carefully, Norman: I... am... lying.
—*Star Trek*, the episode "I, Mudd"

Although John P. A. Ioannidis [1] brings up several good points about over-reliance on formal—yet arbitrary—statistical cutoffs and bias against the reporting of negative results, his claim that most published research findings are false is somewhat paradoxical. Ironically, the truer his premise is, the less likely his conclusions are. He, after all, relies heavily on other studies to support his premise, so if most (i.e., greater than 50%) of his cited studies are themselves false (including the eight of 37 that pertain to his own work), then his argument is automatically on shaky ground. As mentioned in the *PLoS Medicine* Editorial [2], scientific studies don't offer truth, per se. Even when studies appear in the best journals, they offer probabilistic assertions. Ioannidis's statement that "the probability that a research finding is indeed true depends on the prior probability of it being true" [1] is really begging the question; this, after all, is the problem. We cannot know such probabilities a priori, and guessing at such probabilities and/or parameters (as he does in his single nucleotide polymorphism [SNP] association example) surely could not be less biased than any statistical test of significance. The key problem in Ioannidis's positive predictive value (PPV) formula to calculate the post-study probability that a relationship is true (PPV = $[1 - \beta]R/[R - \beta R + \alpha]$, where R is the ratio of true relationships to no relationships) is that one can postulate a near-infinite number of non-relationships. Just extending his SNP example, why assume each SNP acts independently? This is not unreasonable, given that schizophrenia is clearly not inherited in a Mendelian pattern. So rather than 99,990 SNPs not being associated with schizophrenia, we have potentially $99{,}990^n$ not associated, where n is the number of potentially interacting SNPs. As n grows, R becomes very small very quickly, and PPV becomes effectively zero. Taken to the extreme, this would imply that all empirical studies are fruitless. One of the most important factors in moving toward the truth, which was not discussed, is fitting discoveries into a framework. Optimally, if a relationship is true, it should have more than one implication, permitting validation from multiple angles. For example, an SNP causally associated with schizophrenia must affect something on the molecular level, whether genomic, transcriptional, post-transcriptional, translational, or post-translational. In turn, these molecules should interact differently with each other, with other molecules within the cell, within a tissue, and/or with the system as a whole. If Norman, the android from *Star Trek* mentioned in the beginning quote, had been equipped with the capacity to evaluate statements within a framework, he never would have short-circuited as a result of Kirk's paradox. He could have entertained the possibility that either Kirk was lying about Harry or Harry's statement was incomplete (i.e., lying about what?) Similarly, repeatedly re-examining any particular finding to resolve the true/not true paradox via statistical arguments alone can short-circuit our patience. We should instead seek to identify the framework by which implications of the finding can be tested, and I would argue that the more important the finding, the more testable implications it has. ■

Jonathan D. Wren
University of Oklahoma
Norman, Oklahoma, United States of America
E-mail: jonathan.wren@ou.edu

References
1. Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2: e124. DOI: 10.1371/journal.pmed.0020124
2. PLoS Medicine Editors (2005) Minimizing mistakes and embracing uncertainty. PLoS Med 2: e272. DOI: 10.1371/journal.pmed.0020272

Citation: Wren J (2005) Truth, probability, and frameworks. PLoS Med 2(11): e361.

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020361

Author's Reply

I agree with Ian Shrier [1] that, when the effect size is large, adequate power is obtained with small numbers, and it is unnecessary to aim at very large studies. However, most effect sizes probed with statistical testing seem to be small. I also agree that heterogeneity is useful and can offer valuable insights [2]. Sometimes heterogeneity can show us that there are actually two or more research questions, where we thought there was only one [3]. The danger is when heterogeneity is silenced and dismissed in favor of claiming

consistent results and when heterogeneity is exploited to show only the most spectacular results—unfortunately, this is not uncommon.

As Stephen Pauker [4] also points out correctly, it is useful to think about what the post-study odds are that one is aiming for if a study eventually were to get a "positive" result. Some residual uncertainty is unavoidable in any research question, no matter how strong the evidence. We should learn to live with uncertainty. I also agree that often the credibility level is less than 50%, yet decisions still have to be made. I don't see a problem implementing a very safe and very cheap medical intervention, even if the credibility that it is effective is only 20%. However, it is important to understand and acknowledge that this intervention has a credibility of 20%, while another has a credibility of 70%. I have no objection or preference on how exactly this will be calculated and plotted. Likelihood ratios are also a nice equivalent approach to calculate the probabilities or odds.

I agree with Jonathan D. Wren [5] that it is impossible to be 100% certain about the exact pre-study odds of truth for any research, mine included of course. However, I argue that we need to start thinking more seriously and consistently about these pre-study odds. In the single nucleotide polymorphism (SNP) association example, one might argue that 1:10,000 is not the best choice, but I doubt anyone would choose 1:100 [6]. Some fields may, indeed, have a pre-study odds of zero—these are the "null fields" that I discussed [7]. The differences in the range in pre-study odds are huge in current research, and I am afraid that this is almost completely ignored. I also have no objection about the framework concept. It is nice to see multiple lines of evidence converge. In fact "framework" evidence may be used to formulate more accurate pre-study odds. However, we should be cautious about how this framework is interpreted. We need more empirical data on how scientists try to converge various pieces of biological, epidemiological, and clinical information. I suspect that bias to make things fit, even if they don't, is not negligible. ■

John P. A. Ioannidis
University of Ioannina School of Medicine
Ioannina, Greece
E-mail: jioannid@cc.uoi.gr

References
1. Shrier I (2005) Power, reliability, and heterogeneous results. PLoS Med 2: e386. DOI: 10.1371/journal.pmed.0020386
2. Berlin JA (1995) Invited commentary: Benefits of heterogeneity in meta-analysis of data from epidemiologic studies. Am J Epidemiol 142: 383–387.
3. Lau J, Ioannidis JP, Schmid CH (1998) Summing up evidence: One answer is not always enough. Lancet 351: 123–127.
4. Pauker S (2005) The clinical interpretation of research. PLoS Med 2: e395. DOI: 10.1371/journal.pmed.0020395
5. Wren J (2005) Truth, probability, and frameworks. PLoS Med 2: e361. DOI: 10.1371/journal.pmed.0020361
6. Yang Q, Khoury MJ, Friedman JM, Little J, Flanders WD (2005) How many genes underlie the occurrence of common complex diseases in the population? Int J Epidemiol 34: 1129–1137.
7. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2: e124. DOI: 10.1371/journal.pmed.0020124

Citation: Ioannidis JPA (2005) Author's reply. PLoS Med 2(11): e398.

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020398

Did Glycopeptide Use in Animals Result in Hospital Infections of VRE?

Anthony Mudd

As one of the persons involved with the development of avoparcin for farm animals, I have followed the discussion on vancomycin-resistant enterococci (VRE) and its potential transfer from animals to humans over the past decade. What a pity that the authors of this *PLoS Medicine* Policy Forum [1] did not reference a recent review by Wassenaar [2] that comprehensively discussed this topic. In this review, evidence is presented to show that VRE infections in humans have actually increased in the European Union since avoparcin was removed from the market. Other data show that whole-genome typing methods separate clinical VRE strains from animal or nonhospitalized human strains.

The conclusion of the Smith et al. article [1] that a correct decision was made to adopt the EU "precautionary principle" and remove avoparcin from the market is surprising, as this is contrary to the opinion of the independent EU Scientific Committee for Animal Nutrition, and since a quantitative risk analysis, as suggested by the authors, could not conclude a relationship between glycopeptide use in animals and incidence of clinical infection in humans. ■

Anthony Mudd
Southampton, United Kingdom
E-mail: meonconsultants@aol.com

References
1. Smith DL, Dushoff J, Morris G Jr (2005) Agricultural antibiotics and human health. PLoS Med 2: e232. DOI: 10.1371/journal.pmed.0020232
2. Wassenaar TM (2005. Use of antimicrobial agents in veterinary medicine and implications for human health. Crit Rev Microbiol 31: 155–169.

Citation: Mudd A (2005) Did glycopeptide use in animals result in hospital infections of VRE? PLoS Med 2(11): e383.

Competing Interests: AM was a consultant advising the International Federation for Animal Health representing the Global Animal Health companies.

DOI: 10.1371/journal.pmed.0020383

The Need of a Neonatal Preparation for Chagas Disease

Sergio Sosa-Estani, Jose M. Belizan, Fernando Althabe, Aldofo Rubinstein

We have read about the efforts and initiatives related to the design of drugs for parasitic diseases in McKerrow's article [1] with interest and expectation. One of the pressing needs in this area is for a neonatal preparation for Chagas disease.

Satisfactory achievements have been made in Argentina in relation to the transmission of the disease by vectors and through blood transfusion [2,3]. Vertical transmission is now the great challenge in eradicating Chagas disease. Around 800–1,300 neonates infected with *Trypanosoma cruzi* are born every year in our country [4]. Almost 99% of all births occur in hospital, thus allowing the detection of infants born with parasites immediately after birth. The initiation of treatment of these neonates before they and their mothers leave the hospital is a good strategy to obtain high treatment coverage.

failure here, extensively drug-resistant TB may be a possible challenge in Ethiopia. Whether Ethiopia succeeds in the Stop TB Partnership's Global Plan to Stop Tuberculosis, which aims to save 14 million lives between 2006 and 2015 (see http://www.stoptb.org/globalplan), depends on the effectiveness of the national program, infrastructure development, peace, and good governance with sustainable development assistance from donors directed to improving the life condition of the Ethiopian people, so that the population is self-sufficient and confident enough to overcome burning issues like TB.

In conclusion, the study confirms that TB drug delivery, without implementation of anti-poverty programs and more access to public health facilities, is ineffective. ■

Hundie Tesfaye (hundie.tesfaye@fnmotol.cz)
University Hospital in Motol, Charles University
Prague, Czech Republic

References

1. Rothschild BM, Martin LD, Lev G, Bercovier H, Bar-Gal GK, et al. (2001) Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. Clin Infect Dis 33: 305–311.
2. Onyebujoh P, Rook GAW (2004) World Health Organization disease watch. Focus: Tuberculosis. Available: http://www.who.int/tdr/dw/tb2004.htm. Accessed 21 March 2007.
3. World Health Organization (2006) Tuberculosis fact sheet 104. Global and regional incidence. Available: http://www.who.int/mediacentre/factsheets/fs104/en/. Accessed 21 March 2007.
4. Shargie EB, Lindtjørn B (2007) Determinants of treatment adherence among smear-positive pulmonary tuberculosis patients in southern Ethiopia. PLoS Med 4: e37. doi:10.1371/journal.pmed.0040037
5. Centers for Disease Control (CDC) (2006) Emergence of mycobacterium tuberculosis with extensive resistance to second-line drugs—Worldwide, 2000–2004. MMWR Morb Mortal Wkly Rep 55: 301–305.

Citation: Tesfaye H (2007) Adherence to TB treatment in Ethiopia: Why do patients default? PLoS Med 4(4): e165. doi:10.1371/journal.pmed.0040165

Funding: The author received no specific funding for this article.

Competing Interests: The author has declared that no competing interests exist.

Why Most Published Research Findings Are False: Problems in the Analysis

Steven Goodman, Sander Greenland

The article published in *PLoS Medicine* by Ioannidis [1] makes the dramatic claim in the title that "most published research claims are false," and has received extensive attention as a result. The article does provide a useful reminder that the probability of hypotheses depends on much more than just the *p*-value, a point that has been made in the medical literature for at least four decades, and in the statistical literature for decades previous. This topic has renewed importance with the advent of the massive multiple testing often seen in genomics studies.

Unfortunately, while we agree that there are more false claims than many would suspect—based both on poor study design, misinterpretation of *p*-values, and perhaps analytic manipulation—the mathematical argument in the *PLoS Medicine* paper underlying the "proof" of the title's claim has a degree of circularity. As we show in detail in a separately published paper [2], Dr. Ioannidis utilizes a mathematical model that severely diminishes the evidential value of studies—even meta-analyses—such that none can produce more than modest evidence against the null hypothesis, and most are far weaker. This is why, in the offered "proof," the only study types that achieve a posterior probability of 50% or more (large RCTs [randomized controlled trials] and meta-analysis of RCTs) are those to which a prior probability of 50% or more are assigned. So the model employed cannot be considered a proof that most published claims are untrue, but is rather a claim that no study or combination of studies can ever provide convincing evidence.

The two assumptions that produce the above effect are:

1) Calculating the evidential effect only of verdicts of "significance," i.e., $p \leq 0.05$, instead of the actual p-value observed in a study, e.g., $p = 0.001$.

2) Introducing a new "bias" term into the Bayesian calculations, which even at a described "minimal" level (of 10%) has the effect of very dramatically diminishing a study's evidential impact.

In addition to the above problems, the paper claims to have proven something it describes as paradoxical; that the "hotter" an area is (i.e., the more studies published), the more likely studies in that area are to make false claims. We have shown this claim to be erroneous [2]. The mathematical proof offered for this in the *PLoS Medicine* paper shows merely that the more studies published on any subject, the higher the absolute number of false positive (and false negative) studies. It does not show what the papers' graphs and text claim, viz, that the number of false claims will be a higher proportion of the total number of studies published (i.e., that the positive predictive value of each study decreases with increasing number of studies).

The paper offers useful guidance in a number of areas, calling attention to the importance of avoiding all forms of bias, of obtaining more empirical research on the prevalence of various forms of bias, and on the determinants of prior odds of hypotheses. But the claims that the model employed in this paper constitutes a "proof" that most published medical research claims are false, and that research in "hot" areas is most likely to be false, are unfounded. ■

Steven Goodman (sgoodman@jhmi.edu)
Johns Hopkins School of Medicine
Baltimore, Maryland, United States of America

Sander Greenland
University of California Los Angeles
Los Angeles, California, United States of America

References

1. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2: e124. doi:10.1371/journal.pmed.0020124
2. Goodman S, Greenland S (2007) Assessing the unreliability of the medical literature: A response to "Why most published research findings are false". Johns Hopkins University, Department of Biostatistics. Available: http://www.bepress.com/jhubiostat/paper135. Accessed 21 March 2007.

Citation: Goodman S, Greenland S (2007) Why most published research findings are false: Problems in the analysis. PLoS Med 4(4): e168. doi:10.1371/journal.pmed.0040168

Funding: The authors received no specific funding for this article.

Competing Interests: The authors have declared that no competing interests exist.

Funding: The authors received no specific funding for this article.

Competing Interests: The authors have declared that no competing interests exist.

Consent for Genomic Epidemiology in Developing Countries: Added Human Subject Protection Also Needed

Robert Reinhard

The authors deserve thanks for laying out decent principles of communication [1]. But serviceable consent language is insufficient to address all issues of protection. That was the point of recent workshops held by the National Institutes of Health to develop a genome-wide association studies program [2].

Risks associated with personal identification may be incurred if information is subject to code breaking. Legal means are available to compel identification, including across national boundaries. Privacy protections under the Health Insurance Portability and Accountability Act (HIPAA) are subject to exceptions, including for law enforcement, downstream data users, or for other reasons, and are not available internationally. Even with authorization, the complexities associated with a repository may frustrate attempts to achieve meaningful comprehension. Use of data for purposes other than pharmaceutical product development or biomedical interventions would be an abuse resulting perhaps in travel restrictions or discrimination.

For these reasons, safeguards should be added, including:

- Amendments to prevent non-medical health access to personal identification information;
- Restrictions on recruitment of populations especially vulnerable to disclosure risks, such as prisoners or immigrants;
- Prohibitions on disclosure to or use by employers or third-party payors to deny medical coverage, assign differential premium risks, restrict access to therapies, or unfairly discriminate in employment.

Another risk from creation of a genomics repository is the potential for unjust stigmatization (see for example [3]). A workable program would state that the data are appropriate only for limited public health purposes involving product development or professionally derived biomedical intervention, and are insupportable for other use or by political or non-medical entities.

A researcher publishing results based on the genomic data should state affirmatively a boilerplate recognition of the abuse potential for stigmatization. This mechanism could prevent others from the wayward misappropriation of data for purposes other than those intended by professionals. The boilerplate could read:

"Conclusions derived from the genotypic or phenotypic characterization of individuals, groups, or families in this [publication] are meaningful or supportable only for the purpose of biomedical intervention or treatment and are unethical, insupportable, or inappropriate for use in other purposes. Use of the data to support any result of stigmatization, discrimination, or adverse social harm would constitute a misuse or abuse of the data."

To increase the connection of benefits to participants, individuals should be given personal opportunities to receive news reports if they wish and learn of particular clinical trials directed at their characteristics. If the data are to be used in the development of pharmaceutical products, users should also be directed to plan and explain early on how targeted populations may have reasonable access to treatment or therapy if the product is successfully brought to market. These suggestions are consistent with the program outlined by Senator Barack Obama in the Genomics and Personalized Medicine Act of 2006 and Senator Olympia Snowe in the Genetic Information Nondiscrimination Act of 2007 [4,5].

Improved consent: Yes, but linked to and inseparable from strong protections and added benefits for participants. ■

Robert Reinhard (rreinhard@mofo.com)
San Francisco Vaccine Trials Unit
San Francisco, California, United States of America

References

1. Chokshi DA, Thera MA, Parker M, Diakite M, Makani J, et al. (2007) Valid consent for genomic epidemiology in developing countries. PLoS Med 4: e95. doi:10.1371/journal.pmed.0040095
2. National Institutes of Health (2006) Request for information (RFI): Proposed policy for sharing of data obtained in NIH supported or conducted genome-wide association studies (GWAS). Available: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-06-094.html. Accessed 24 May 2007.
3. Council for International Organizations of Medical Sciences (2006) Special ethical considerations for epidemiological research.
4. US Congress (2006) The Genomics and Personalized Medicine Act of 2006. S. 3822, 109th Congress, 2d session. Available: http://www.govtrack.us/congress/billtext.xpd?bill=s109-3822. Accessed 24 May 2007.
5. US Congress (2007) Genetic Information Nondiscrimination Act of 2007. S. 358, 110th Congress, 1st Session. Available: http://www.govtrack.us/congress/bill.xpd?bill=s110-358. Accessed 24 May 2007.

Citation: Reinhard R (2007) Consent for Genomic Epidemiology in Developing Countries: Added human subject protection also needed. PLoS Med 4(6): e214. doi:10.1371/journal.pmed.0040214

Funding: The author received no specific funding for this article.

Competing Interests: The author has declared that no competing interests exist.

Why Most Published Research Findings Are False: Author's Reply to Goodman and Greenland

John P. A. Ioannidis

I thank Goodman and Greenland for their interesting comments [1] on my article [2]. Our methods and results are practically identical. However, some of my arguments are misrepresented:

1. I did not "claim that no study or combination of studies can ever provide convincing evidence." In the illustrative examples (Table 4), there is a wide credibility gradient (0.1% to 85%) for different research designs and settings.

2. I did not assume that all significant *p*-values are around 0.05. Tables 1–3 and the respective positive predictive value (PPV) equations can use any *p*-value (alpha). Nevertheless, the $p = 0.05$ threshold is unfortunately entrenched in many scientific fields. Almost half of the "positive" findings in

recent observational studies have *p*-values of 0.01–0.05 [3,4]; most "positive" trials and meta-analyses also have modest *p*-values.

3. I provided equations for calculating the credibility of research findings with or without bias. Even without any bias, PPV probably remains below 0.50 for most non-randomized, non-large-scale circumstances. Large trials and meta-analyses represent a minority of the literature.

4. Figure 1 shows that bias can indeed make a difference. The proposed modeling has an additional useful feature: As type I and II errors decrease, PPV(max) = 1 − [u/(R + u)], meaning that to allow a research finding to become more than 50% credible, we must first reduce bias at least below the pre-study odds of truth (u less than R). Numerous studies demonstrate the strong presence of bias across research designs: indicative reference lists appear in [5–7]. We should understand bias and minimize it, not ignore it.

5. "Hot fields": Table 3 and Figure 2 present "the probability that at least one study, among several done on the same question, claims a statistically significant research finding." They are not erroneous. Fields with many furtive competing teams may espouse significance-chasing behaviors, selectively highlighting "positive" results. Conversely, having many teams with transparent availability of all results and integration of data across teams leads to genuine progress. We need replication, not just discovery [5].

6. The claim by two leading Bayesian methodologists that a Bayesian approach is somewhat circular and questionable contradicts Greenland's own writings: "One misconception (of many) about Bayesian analyses is that prior distributions introduce assumptions that are more questionable than assumptions made by frequentist methods" [8].

7. Empirical data on the refutation rates for various research designs agree with the estimates obtained in the proposed modeling [9], not with estimates ignoring bias. Additional empirical research on these fronts would be very useful.

Scientific investigation is the noblest pursuit. I think we can improve the respect of the public for researchers by showing how difficult success is. Confidence in the research enterprise is probably undermined primarily when we claim that discoveries are more certain than they really are, and then the public, scientists, and patients suffer the painful refutations. ■

John P. A. Ioannidis (jioannid@cc.uoi.gr)
University of Ioannina School of Medicine
Ioannina, Epirus, Greece

References

1. Goodman S, Greenland S (2007) Why most published research findings are false: Problems in the analysis. PLoS Med 4: e168. doi:10.1371/journal.pmed.0040168
2. Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2: e124. doi:10.1371/journal.pmed.0020124
3. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, et al. (2004) Issues in the reporting of epidemiological studies: A survey of recent practice. BMJ 329: 883.
4. Kavvoura FK, Liberopoulos G, Ioannidis JP (2007) Selection in reported epidemiological risks: An empirical assessment. PLoS Med 4: e79. doi:10.1371/journal.pmed.0040079
5. Ioannidis JP (2006) Evolution and translation of research findings: From bench to where? PLoS Clin Trials 1: e36. doi:10.1371/journal.pctr.0010036
6. Gluud LL (2006) Bias in clinical intervention research. Am J Epidemiol 163: 493–501.
7. The Cochrane Collaboration (2007) Cochrane methodology register. Available: http://www3.cochrane.org/access_data/cmr/accessDB_cmr.asp. Accessed 23 May 2007.
8. Greenland S (2006) Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol 35: 765–775.
9. Ioannidis JP (2005) Contradicted and initially stronger effects in highly cited clinical research. JAMA 294: 218–228.

Citation: Ioannidis JPA (2007) Why Most Published Research Findings Are False: Author's reply to Goodman and Greenland. PLoS Med 4(6): e215. doi:10.1371/journal.pmed.0040215

Funding: The author received no specific funding for this article.

Competing Interests: The author has declared that no competing interests exist.

Biomedical Journals and Global Poverty: Is HINARI a Step Backwards?

Javier Villafuerte-Gálvez, Walter H. Curioso, Oscar Gayoso

Much has been written about how open access to biomedical journals is vital for researchers in developing countries [1], but so much more needs to be done.

Our experience in Peru with the Health InterNetwork Access to Research Initiative (HINARI), an initiative managed by the World Health Organization that helps promote access to scientific information by providing free (or low cost) online access to major science journals, is not as accessible as hoped for and, in fact, is getting worse. When HINARI launched in 2003, it provided access to more than 2,300 major journals in biomedical and related social sciences [2].

In April 2007, we conducted a review of the first 150 science journals available through HINARI with the highest impact factors on the Science Citation Index [3]. We excluded open-access journals and journals that make online access free to low-income countries (e.g., *The New England Journal of Medicine*, British Medical Journal Publishing Group). We could not access any of the top five journals from major publishers such as Nature and Elsevier-Science Direct. In other words, from the Nature Publishing Group we had no access to *Nature Reviews Cancer*, *Nature Reviews Immunology*, *Nature Reviews Molecular Cell Biology*, *Nature*, or *Nature Medicine*, and from Elsevier ScienceDirect we had no access to *Cell*, *Cancer Cell*, *Current Opinion in Cell Biology*, *Immunity*, or *Molecular Cell*. In addition, we could not access any of the first-level journals from Blackwell, Oxford Press University, Lippincott Williams and Wilkins, or Wiley and Sons. In 2003, all these journals were available.

Our findings support comments received from users over the last 8–10 months at the main library at Universidad Peruana Cayetano Heredia (Oscar Gayoso, personal communication). Students and faculty could not get access to biomedical journals from Nature, Elsevier-Science Direct, Blackwell, Oxford Press University, Springer Science, Lippincott Williams and Wilkins, or Wiley and Sons through HINARI. The collections of journals from the above-mentioned publishers together represent approximately 57% (2,118 of 3,741) of journals that were supposed to be accessible through HINARI, while the remaining 43% accessible were largely composed of open-access journals or journals that make online access free to low-income countries.

Moreover, we have found a significant decrease in the number of users accessing HINARI at our institution. For example, the number of HINARI users has decreased from