Evidence and replication

Large scale evidence and replication: insights from rheumatology and beyond

J P A Ioannidis

Many biomedical and research findings fail to be replicated and validated

Increasing concern has been expressed that biomedical research findings many times fail to be replicated and validated. This may occur across the spectrum from basic research to clinical applications. Small studies, poor design, the play of chance, exaggeration in early claims, data dredging, selective publication and reporting of information, strong expert opinions, biases, and financial or other conflicts of interest all probably combine to create uncertainty about whether research findings will stand the test of time and independent replication. Some research findings or beliefs are not even questioned and no effort is ever made to challenge them, but this does not mean that they are true. The recent advent of molecular medicine has only increased the complexity and number of questions that may be asked. It is unclear whether the information derived from the various fascinating discovery-driven approaches, massive as it is, is more likely to be possible to replicate. The ratio of sample size to scientific question asked has been decreased to infinitesimal levels, and uncertainty remains as high as ever.

Small sample sizes: empirical data

For rheumatology and autoimmune diseases, the danger is as great as for any other medical discipline. Much of the replication difficulties may have to deal with the fact that claims are made or refuted based on limited data in single studies. The median number of patients in randomised controlled trials in systemic sclerosis is only 28,1 exactly of them asking anywhere between one and 90% of the discoveries are false positive claims.4 The high rate of false discoveries has recently prompted several journals, including Arthritis and Rheumatism,9 to recommend replication of genetic associations in independent populations in the same publication. For randomised trials, theoretically the “success rate” may be better, but conflicts of interest are strong,10 so misleading inferences are probably not uncommon at all.

Possibilities for large scale evidence

This situation suggests that it is essential to generate large scale evidence on clinical and research questions of interest, regardless of how strong the data seem on preliminary small studies. It is also very important not to accept evidence, no matter how substantial, at face value, but to scrutinise it for potential errors and biases.

“Even apparently good evidence should be closely examined for possible errors and biases”

There are several different approaches to obtaining large scale evidence. Single centres can occasionally generate large scale information, provided that massive data can be collected meticulously over many years. For example, we recently managed to collect data on almost 800 patients with primary Sjögren’s syndrome followed up at a single centre over a period of 20 years and the evidence was strong enough to suggest that primary Sjögren’s syndrome is probably two different diseases, one with a high risk of lymphoproliferative disease and death and, another, far more common, with no different risk from the general population.11 However, in the vast majority of settings, single centre experiences would be fragmented snapshots. Multicentre collaborations should be encouraged. For example, such multicentre collaborations have recently been able to establish databases on 1000 well characterised...
patients with systemic lupus erythematosus and 1000 patients with antiphospholipid syndrome.

Once data are obtained from different centres and different settings, it is important to try to dissect and estimate the variability that exists between the various pieces of information. In a recent collaborative project on systemic sclerosis, we accumulated information on 3311 patients and 19 990 patient-years of follow up. The most interesting finding was the large diversity that existed between the different cohorts contributing data. In some cohorts we could identify patients with a low risk profile who had a risk of death similar to that of the general population, while in other cohorts with stronger referral patterns, patients with a similar “low risk profile” actually had a very high risk of death. Even for well established diseases with clear criteria, different centres may use different definitions of the diseases/syndromes and evidence may be used to improve the standardisation of multicentre evidence would be indispensable.

For several autoimmune diseases, large scale evidence may be used to improve the definitions of the diseases/syndromes themselves, and in most cases large scale multicentre evidence would be indispensable to improve the standardisation of disease outcome.

**“BIG” SCIENCE IS NOT NECESSARILY “GOOD” SCIENCE**

We should be cautious and realise that large sample sizes are not necessarily synonymous with high quality research and offer no guarantee that the research is scientifically or clinically relevant. Some randomised trials funded by industry manage to achieve very large sample sizes.

“Large sample sizes do not garanttee that the results are reliable”

This primarily serves the purpose of being certain that statistically significant results will be achieved eventually, even if the absolute treatment effects for the selected outcomes are actually trivial from a clinical and scientific viewpoint. Sample size seen in isolation is not a reliable marker of high scientific standards.

**EVIDENCE IS AN EVOLVING PUZZLE**

Prospective large scale evidence is preferable to collation of already available retrospective data. However, for most research questions in rheumatology and beyond, it is unlikely that we will ever have the single, perfect, definitive, prospective study. Even if we do, it is likely that other prospective and retrospective studies may also be conducted on the same or similar questions. Thus it is becoming increasingly important to view evidence in its totality and to examine its continuous evolution over time as new data accumulate. This suggests that meta-analytic methodologies are likely to have increasing impact in the future.

Meta-analysis, the rigorous and systematic quantitative synthesis of information on the same question, was initially promoted as a tool for pooling data and improving power. Although this is useful, simple pooling does not suffice. As described above, errors and biases abound and naïve meta-analysis may do more harm than good by simply summing up erroneous information and leading to erroneous confidence in spuriously “precise” results.

We need more emphasis on methods and diagnostics that would examine rigorously the accumulating evidence for the presence of heterogeneity, bias, or both. Meta-analysis provides a unique opportunity for a critical appraisal of the quality of the data, the potential gaps, sources of diversity, and sources of errors. Feedback from the meta-analyses on these identified sources of diversity and problems may also be used to shape and correct the future research agenda.

We should pay more attention to the whole picture and the problems that it reveals. For example, 15 years ago, it was clear that we already had large scale evidence, although fragmented, on nonsteroidal anti-inflammatory agents in rheumatoid arthritis, and that several problems were commonly encountered in the design and reporting of these trials. Fifteen years later we continue to perform a lot of trials on this area, but we continue also making many of the same mistakes, while there is also the potential for new types of error as the standard requirements increase.

**CONCLUSION**

In all, this editorial should be read as a plea for evidence. Solid evidence requires large scale and good quality information. It also requires a thorough appreciation of the whole picture and this may be constantly changing over time as more data accumulate. The systematic and critical monitoring of the big picture can be a very important research endeavour on its own and we have a lot to learn from it.


Correspondence to: Dr J P Ioannidis, Ioannina School of Medicine, Ioannina, Greece, jpioannidis@cc.uoi.gr

Accepted for publication 14 September 2004

**REFERENCES**


20. Gotzsche PC. Reporting of outcomes in arthritis trials measured on ordinal and interval scales is inadequate in relation to meta-analysis. Ann Rheum Dis 2001;60:349–52

www.annrheumdis.com
Large scale evidence and replication: insights from rheumatology and beyond

J P A Ioannidis

Ann Rheum Dis 2005 64: 345-346 originally published online September 30, 2004
doi: 10.1136/ard.2004.027979

Updated information and services can be found at:
http://ard.bmj.com/content/64/3/345

These include:

References
This article cites 19 articles, 3 of which you can access for free at:
http://ard.bmj.com/content/64/3/345#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

- Immunology (including allergy) (5144)
- Connective tissue disease (4253)
- Musculoskeletal syndromes (4951)
- Rheumatoid arthritis (3258)
- Systemic lupus erythematosus (571)
- Degenerative joint disease (4641)
- Osteoarthritis (931)
- Epidemiology (1367)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/