Written Testimony of
Dr. Bruce M. Alberts
Editor-in-Chief, *Science* magazine;
Professor Emeritus, UCSF
Before the Subcommittee on Research,
Committee on Science, Space, and Technology
U.S. House of Representatives
Hearing on
"Scientific Integrity and Transparency"
March 5, 2013

Mr. Chairman, Ranking Member Lipinski, and Members of the Subcommittee, my name is Bruce Alberts and I currently serve as the Editor-in-Chief of *Science* magazine. I thank you for the opportunity to speak to you today on this important topic for the future of science and the United States.

Science magazine is a leading weekly science journal (100,000 subscriptions) published by the American Association for the Advancement of Science (AAAS). I am a biochemist and cell biologist whose major research contributions have concerned the mechanism of DNA replication, which is the process that duplicates chromosomes before a cell divides. A Professor Emeritus in the School of Medicine at the University of California, San Francisco (UCSF), I have recently served as one of the first three U. S. Science Envoys, appointed by Secretary of State Hillary Clinton. My previous positions include: full-time president of the National Academy of Sciences (1993-2005), president of the American Society for Cell Biology, and chairman of the Department of Biochemistry and Biophysics at UCSF. I am a member of the National Academy of Sciences, and a foreign member of the Royal Society (UK), the Indian National Science Academy, and the academies of several other nations.

As I have written in many editorials for *Science*, the strength of US science and technology (S & T) has been, and will long be, critical for our position as the leading nation of the world. It underlies both our economic success and our military dominance. In recent years, nations like China have focused intensely on strengthening their own S & T as they increasingly challenge our leadership position. Critical to maintaining the position of the US in the world will be both the amount and the quality of our long-term fundamental research in science, engineering, and medicine. The National Academies outlines the value of basic science in a series of twenty pamphlets on such research that has led in the past to breakthroughs with great human and economic benefit. Three of the 20 highlighted examples were the global positioning system, modern communications, and the antiviral therapy for AIDS. (See www.beyonddiscovery.org.) Exactly how future advances in our fundamental understanding of the universe will lead to such benefits can never be predicted in advance. Nevertheless, based on past experience, we can confidently expect striking breakthroughs to emerge from such research that are completely unimaginable now.

Although the subject of basic science funding is not a focus of today's hearing, the House Science, Space, and Technology Committee has long emphasized the critical importance of our investments in America's future through governmental support of fundamental, long-term research. This is an investment that has remained stagnant in the U.S., while other nations are increasing their research intensity at an alarming pace. According to the AAAS, this type of investment will have decreased in the U.S. as a percent of GDP from 1.25 percent in 1985 to 0.87 percent of GDP in 2013. And in a ranking of total R&D spending as a share of GDP, America came in tenth in 2011, whereas we were sixth in 2001. The sequester will now make the situation considerably worse.

But this Hearing is entirely focused on the quality of U.S. scientific research and how we might improve it. I shall now proceed to address the specific questions posed.

Why is the integrity of scientific results and data sharing so important for both the scientific community and the general public?

Science is a remarkable community endeavor, in which a reliable body of knowledge about how the world works, called Science (with an upper case S), is built up over time from the many small bits of science (with a lower case s) that is carried out by large numbers of individual scientists. The rules established for individual scientists that make it all work demand that – in return for being given the privilege of publishing any particular research finding – each scientist must provide access to the methods that he or she has used, as well as to the data, so that any one else in the world can try to repeat the work to either confirm or deny what the first scientist has claimed. Once thereby, confirmed, new knowledge is developed by building on this knowledge in novel ways, through the work of many other scientists. Integrity and data sharing are crucial, because scientists are constantly relying on the discoveries of others as they carry on their own research. Without both the integrity of scientific results and data sharing, Science cannot develop from science.

Why does the public have such a strong interest in this issue? It is because of the enormous benefits that the public derives from Science, as explained in each of the 20 case studies that I described earlier entitled "Beyond Discovery: the Path from Research to Human Benefit" (www.beyonddiscovery.org). Such benefits are precisely why governments invest so heavily in supporting scientific research for the public good. Thus the scientific community places great emphasis on promoting the highest scientific ethics, using aids such as the freely available publication from the National Academies "On Being a Scientist: A Guide to Responsible Conduct in Research" to help imbue the needed scientific values in the next generation of scientists.[1]

What factors have contributed toward a scientific culture where unreliable results are being published and data sharing is difficult?

I begin with the data sharing issue, which is the easier half of this question to answer. The others testifying today will address ways to make the scientific data that is produced by one scientist more widely accessible and reusable for other scientists. This is an important issue for all fields of science, and *Science* magazine has long strongly supported such efforts. In early 2011, we published a large special issue of the magazine entitled "Dealing with Data" that contained 14 articles on its different aspects, in fields from astronomy to genomics. In our editorial for that issue, entitled "Making Data Maximally Available," we stressed that "*Science's* policy for some time has been that "all data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*" (see www.sciencemag.org/site/feature/contribinfo/)." And we announced a new policy that extended this requirement "to include computer codes involved in the creation or analysis of data." [2]

In general, we feel that data availability has increased dramatically in recent decades and that more data is available now than ever before. Standards are firmer and the community norms have improved. Problems of course remain, in part due to the massive amounts of data that can now be rapidly collected. Many of these were highlighted in our "Dealing with Data" special issue.

The main current challenges that we see with regard to data sharing are:

1) Developing standards on what data to keep, inasmuch as some scientists are collecting terabytes (TB) of data daily.

2) Developing community standards for how to organize and describe the data that is kept and where exactly to deposit it.

3) Guaranteeing funding for public databases long term, so that the community and journals like ours can rely on them.

4) Developing standards for how to deal with huge datasets that have to be housed locally, and providing protocols to access the data.

5) Developing tools for interacting with the large datasets that are now increasingly provided as supplemental online information in journal publications like ours.

The other half of the Committee's question is more difficult to address. At least in large part, I believe that the concern about unreliable results being published reflects reports stating that many of the results in the field of "translational medical research" cannot be reproduced by other scientists.[3] Much of the research that cannot be reproduced aims at identifying the specific protein targets that could be useful to pharmaceutical and biotechnology companies seeking to develop new drugs.

Even though I have been a faculty member in the School of Medicine at UCSF since 1976, I have never carried out this type of research myself. Instead, like many of my colleagues, I have pursued basic mechanistic studies aimed at understanding biological systems at the molecular level. In preparation for this testimony, I have therefore spoken to top scientists at Genentech, a very successful biotechnology company that frequently makes use of results published in the scientific literature for their own research into potential drug targets. In general, they agree with the conclusions concerning drug target reproducibility published by the Bayer HeathCare scientists in reference 3. One of the groups of Genentech scientists whom I consulted was Dr. Frederic de Sauvage, who pointed me to a paper that he published in 2008 that refuted the results of 7 earlier publications in very prestigious journals (references 1 to 7 in his paper). [4]

From this and many other discussions that I have had on this issue, I have reached a few tentative conclusions.

1) The first is that the scientific standards are lower in some subfields of science than others. For example, I am told that many published papers in medically related fields have not been officially retracted by either the journal or the authors, even though the authors have agreed with those unable to reproduce their results that their original publication is wrong. We need to develop a value system where simply "moving on from one's mistakes without publicly acknowledging them" severely damages, rather than protects, a scientific reputation.

2) Human cells are incredibly complex. Because their behavior is determined by huge networks of interlocked signaling pathways, an off-target effect -- one that is due to affecting a protein other than the intended one -- will often mimic the expected effect for a hit on the desired drug target. Every scientist should be trained to be highly suspicious about his or her own results. But a scientist whose career advancement requires finding a drug target may fail to carry out all of the many controls needed to avoid reaching a false conclusion. And the pressures on and incentives for a young researcher whose focus is finding a potential drug target can make it difficult to avoid inadvertent data selection. .

3) We are currently overemphasizing research directly aimed at finding drugs at the expense of the high quality discovery-driven basic research that is urgently needed to improve the search for disease treatments. As elegantly pointed out in a recent editorial in *Science* by Dr. Huda Zoghbi, a leading researcher in translational medicine and a member of the U.S. National Academy of Sciences: [5]

> "Science, like most human endeavors, is susceptible to fads and fashions driven by money and status; and today many highly qualified basic scientists feel compelled to jump on the "translational medicine" bandwagon. For quite some time, it has been apparent that biomedical research in the United States is more likely to get funded if it is tied to a practical outcome, such as a step toward a cure for some disorder. There is no doubt that such targeted and in-depth disease-oriented research is sorely needed. But it is at least as important to support investigators dedicated to discovery-driven basic research." She then goes on to observe that the "task of translational research is not unlike the act of translating a book from one language into another. Fluency in both languages is a given; beyond that, there must be a talent, a feel, for those concepts unique to one language or culture that cannot be directly translated

but must somehow still be conveyed. The challenge in translational medicine is that scientists are trying to translate a text with the sophistication and depth of Shakespeare using a first-grader's vocabulary and experience, because our knowledge about the functions of most pathways in various cell types, during different developmental stages, and under normal physiological conditions, is still rudimentary and piecemeal."

What issues must be considered when promoting the publication and responsible sharing of data? From your experience, what are some models that have worked?

Scientific journals like ours have an important role to play in enforcing the responsible sharing of data. As stated previously, when a scientist publishes research with us, he or she must agree that "all data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*." There have only been rare times when we have had to reinforce this provision with an author. In the early 1990s, several journals, including *Science*, *Nature*, and *Cell* joined together to require X-ray crystallographic data to be made publicly available in a shared database immediately upon publication (some of the scientists involved instead wanted a 1-year moratorium on this release). All genomic data must meet the same type of standard before being published. In addition, in the late 1990's, many journals started to publish data supplements and require electronic data deposition with the journal. A continuing problem is presented by huge datasets in fields where there is no public database for deposition. Here *Science* has had an archival agreement with authors. (They have to house large datasets for 5 years and we get an escrow copy). But such data storage must be paid for by the journal and the cost is a perpetual one; thus, it is not clear how long this type of service can be maintained.

A different critical need that all journals should enforce is the clear and complete presentation in each publication of all of the materials and methods that were used in the research. This goal has become much easier to attain due to the Internet, because the limited space in the printed journal is now routinely supplemented by online supplementary material that is made readily available electronically.

In December 2011, *Science* published a special issue entitled "Data Replication and Reproducibility." This is a topic that we shall return to again in the future. As have other journals, *Science* has on occasion been fooled into publishing articles that contain data that was fabricated by one or more of the authors. As soon as possible after either an honest error or a fraud is detected, the retracted papers are specifically highlighted as incorrect, so that anyone accessing the paper on our website will know that it is wrong. Although ideally a paper will be publicly retracted by its authors, the Editor-in-Chief has retracted incorrect papers in the absence of such consent.

To help protect against both data selection (scientists fooling themselves) as well as against the much rarer intentional fabrication of data, *Science* has initiated a policy to help senior scientists enforce standards in their own laboratories. As I announced in an editorial on January 1, 2010 entitled "Promoting Scientific Standards": [6]

> “*Science* will require that the senior author for each laboratory or group confirm that he or she has personally reviewed the original data generated by that unit, ascertaining that the data selected for publication in specific figures and tables have been appropriately presented. Thus, for example, a researcher who prepares a digitally processed figure displaying an assortment of electrophoretic gel separations will need to present all of the original gel data to a specified senior author, who must certify that this has been done when the manuscript is returned for revision. In this way, *Science* aims to identify a few senior authors who collectively take responsibility for all of the data presented in each published paper. Traditionally, a single individual has been asked to accept this responsibility. But the former requirement has become increasingly unrealistic, considering that a large fraction of publications now contain contributions from groups with very different expertise—and that half of the papers published in 2009 by *Science* had authors from more than one nation.”

I believe that there is more that can and should be done to enforce scientific standards by the community. For example, I strongly favor the proposal that the biosketches routinely used to help evaluate an individual researcher for research support, appointments, and promotions be limited to a small number of publications, for each of which both the significance and the contribution of the individual must be carefully described. It is time to stop allowing long lists of publications to be presented, many of which (in some fields) may have contained major errors, despite having been cited extensively in the literature.

I also believe that new experiments are in order, aimed at creating a much lower barrier for reporting any serious effort to reproduce results that has failed, and insuring that such information becomes attached directly to the original publication in a way that cannot easily be missed.

To summarize, improving the quality of scientific publications will require an on-going effort by many different players in the scientific community. Scientific journals like ours will need to play leadership roles in enforcing standards. Checklists are beginning to be developed by the community to help both scientists and journals guard against the most common errors in research in selected fields like drug target development. Funding agencies can help by facilitating and rewarding the publication of failures to replicate important published results, as well as by changing the way that the biosketches in grant submissions are presented and evaluated. And research institutes and universities should place more emphasis on short courses that teach research methodology, ethics, and important technical skills such as how to avoid statistical errors to all of their research trainees. In fact, I myself will be co-teaching such an intensive two-week “minicourse” to PhD students at UCSF this coming May.

REFERENCES

1. Committee on Science, Engineering, and Public Policy, *On Being a Scientist: A Guide to Responsible Conduct in Research.* The National Academies Press, 2009
2. B. Hanson, A. Sugden, and B. Alberts, Making Data Maximally Available. *Science* 331: 649 (2011).
3. F. Prinz, T. Schlange, and K. Asadullah, Believe it or not: How much can we rely on published data on potential drug targets? *Nature Rev. Drug Discov.* 10: 712 (2011).
4. R. L. Yauch, et al., A paracrine requirement for hedgehog signalling in cancer. *Nature* 455:406 (2008).
5. H. Zoghbi, The Basics of Translation. *Science* 339:250 (2013).
6. B. Alberts, Promoting Scientific Standards. *Science* 327:12 (2010).

Witness Biography

BRUCE ALBERTS   EDITOR-IN-CHIEF, SCIENCE MAGAZINE; PROFESSOR EMERITUS, UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Bruce Alberts, a prominent biochemist with a strong commitment to the improvement of science and mathematics education, serves as Editor-in-Chief of *Science* and served as one of President Obama's first three Science Envoys. Alberts is also Professor Emeritus in the Department of Biochemistry and Biophysics at the University of California, San Francisco, to which he returned after serving two six-year terms as the president of the National Academy of Sciences (NAS).

During his tenure at the NAS, Alberts was instrumental in developing the landmark National Science Education standards that have been implemented in school systems nationwide. The type of "science as inquiry" teaching we need, says Alberts, emphasizes "logical, hands-on problem solving, and it insists on having evidence for claims that can be confirmed by others. It requires work in cooperative groups, where those with different types of talents can discover them – developing self confidence and an ability to communicate effectively with others."

Alberts is also noted as one of the original authors of *The Molecular Biology of the Cell*, a preeminent textbook in the field now in its fifth edition. For the period 2000 to 2009, he served as the co-chair of the InterAcademy Council, an organization in Amsterdam governed by the presidents of 15 national academies of sciences and that was established to provide scientific advice to the world.

Committed in his international work to the promotion of the "creativity, openness and tolerance that are inherent to science," Alberts believes that "scientists all around the world must now band together to help create more rational, scientifically-based societies that find dogmatism intolerable."

Widely recognized for his work in the fields of biochemistry and molecular biology, Alberts has earned many honors and awards, including 16 honorary degrees. He currently serves on the advisory boards of more than 20 non-profit institutions, including the Gordon and Betty Moore Foundation.