resulted in a significant reduction in chronic GVHD without an increased risk of relapse or infection. It is important to know that antilymphocyte globulin is produced in rabbits after vaccination with the human Jurkat T-cell line, whereas antithymocyte globulin is produced after vaccination with human thymocytes. Because of the different immunologic properties of these distinct preparations and the lack of reliable comparative studies, the different brands and doses are not interchangeable and our results with antilymphocyte globulin may not be generalizable to antithymocyte globulin.[3,4]

Nicolaus Kröger, M.D.
University Medical Center Hamburg
Hamburg, Germany
nkroeger@uke.de

Carlos Solano, M.D.
Hospital Clinico Universitario
Valencia, Spain

Francesca Bonifazi, M.D.
S. Orsola-Malpighi University Hospital
Bologna, Italy

Since publication of their article, the authors report no further potential conflict of interest.

1. Bonifazi F, Bandini G, Arpinati M, et al. Intensification of GVHD prophylaxis with low-dose ATG-F before allogeneic PBSC transplantation from HLA-identical siblings in adult patients with hematological malignancies: results from a retrospective analysis. Bone Marrow Transplant 2012;47:1105-11.
2. Wolschke C, Zabelina T, Ayuk F, et al. Effective prevention of GVHD using in vivo T-cell depletion with anti-lymphocyte globulin in HLA-identical or -mismatched sibling peripheral blood stem cell transplantation. Bone Marrow Transplant 2014;49:126-30.
3. Ayuk F, Diyachenko G, Zabelina T, et al. Comparison of two doses of antithymocyte globulin in patients undergoing matched unrelated donor allogeneic stem cell transplantation. Biol Blood Marrow Transplant 2008;14:913-9.
4. Finke J, Bethge WA, Schmoor C, et al. Standard graft-versus-host disease prophylaxis with or without anti-T-cell globulin in haematopoietic cell transplantation from matched unrelated donors: a randomised, open-label, multicentre phase 3 trial. Lancet Oncol 2009;10:855-64.

DOI: 10.1056/NEJMc1601364

More on Data Sharing

TO THE EDITOR: For all the understandable uproar over the term "research parasites" — an inflammatory term that gives short shrift to how open data changed our understanding of Tamiflu, Paxil, and other treatments — those of us who support increased data sharing should realize that Drazen and Longo[1,2] were giving voice to an opinion that many researchers privately hold. After all, it is only human nature that some feel wary of a policy that seems to require them to do extra work that other people will then use for their own academic advancement.

The best way to create a world with more data sharing is to hear out these concerns fairly and figure out how to address them. For example, tenure committees and National Institutes of Health funding reviews should give abundant credit to anyone who originates a data set that other scientists find useful. If data sharing is in the self-interest of whoever collected the data, data sharing as a policy will be on better footing.

Stuart Buck, J.D., Ph.D.
Laura and John Arnold Foundation
Houston, TX
stuartbuck@gmail.com

No potential conflict of interest relevant to this letter was reported.

1. Longo DL, Drazen JM. Data sharing. N Engl J Med 2016;374:276-7.
2. Drazen JM. Data sharing and the *Journal*. N Engl J Med 2016;374(19):e24.

DOI: 10.1056/NEJMc1602586

TO THE EDITOR: Scientific discoveries are public goods, but public goods are typically underproduced in a market economy because producers are not adequately rewarded. The norms of academic science resolve this social dilemma by rewarding scientists with recognition — which translates into career benefits — for making their discoveries public.[1]

In science, most recognition accrues to the first to make a discovery.[2] Scientists thus need to report their discoveries rapidly to avoid being scooped. Society benefits because other teams can then immediately build upon that work.

Longo and Drazen question whether, outside a collaborative relationship, researchers should be permitted to independently analyze data collected by others. But the alternative would allow those who generate data to grab recognition for a discovery and still restrict access to those data. That alternative would have massive unintended consequences. When data can be withheld, re-

The New England Journal of Medicine
Downloaded from nejm.org on May 17, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

searchers can have their cake, by hoarding their data, and eat it, too, by claiming public credit. Should such behavior become widespread, production of public goods would diminish and the pace of discovery would slow.

Carl T. Bergstrom, Ph.D.

University of Washington
Seattle, WA
cbergst@u.washington.edu

No potential conflict of interest relevant to this letter was reported.

1. Stephan PE. The economics of science. J Econ Lit 1996;Sept 1:1199-235.
2. Merton RK. Priorities in scientific discovery: a chapter in the sociology of science. Am Sociol Rev 1957;22:635-59.

DOI: 10.1056/NEJMc1602586

TO THE EDITOR: In their editorials, Longo and Drazen describe the moral imperative of data sharing and the concerns of clinical trialists regarding potential challenges and pitfalls of data reuse by so-called research parasites. They suggested that secondary research should be limited to that involving direct collaboration with primary researchers. While we agree that collaborative secondary research is important, we strongly believe that when it is done right, broader data sharing and reuse can benefit all stakeholders — patients, secondary researchers, and primary researchers. Indeed, the impact of quality research should be enhanced, not diminished, by responsible data sharing and reuse. Rather than discouraging activities that promise sorely needed insights to benefit human health, we should be encouraging and supporting mechanisms that enable responsible secondary research. Examples of positive solutions to the concerns described by Longo and Drazen include improving ethical oversight and peer-review processes to assess the reliability of secondary science, storing metadata with primary data sets to ensure proper contextualization for further analyses, and expanding new metrics for tracking and appropriately crediting researchers whose data sets lead to downstream discoveries.[1]

Peter J. Embi, M.D.

Ohio State University
Columbus, OH
embi.1@osu.edu

Jessica D. Tenenbaum, Ph.D.

Duke University
Durham, NC

No potential conflict of interest relevant to this letter was reported.

1. Kratz JE, Strasser C. Comment: making data count. Sci Data 2015;2:150039.

DOI: 10.1056/NEJMc1602586

THE AUTHORS REPLY: These letters and others we received, as well as comments that dotted the Internet, reflect a broad range of opinions about the implementation of data-sharing policies. The letters affirm that most people recognize the potential value of studying existing data, gathered in the setting of clinical trials, for new findings; the problem is how to make that happen in a way that respects the needs and norms of the various communities involved.

The point of our editorial was that among the models for sharing data from clinical trials, collaboration between the primary data gatherers and the reanalysts is one that respects the wishes of patients who put themselves at risk for the data to be gathered, reflects the effort of the people who accrued the data, and provides a convenient format for reanalysts to work with available data. We hope that those who gathered the data will recognize the benefit of having others reproduce, confirm, and extend their results. Moreover, we think that the likelihood that others will be examining archived data sets will lead to better data management and curation, thus making it easier for others to benefit from the fruits of their labors.

In the circumstance in which data are not used in collaboration with the data gatherers, we think that reanalysts should, at a minimum, demonstrate that they can obtain the same primary findings that the data collectors published. This will help the reanalysts to avoid making misleading statements that derive from an inadequate understanding of the clinical trial. In this circumstance, as noted in our editorial, a mechanism to give full academic credit to those who gathered the data needs to be developed.

We believe that sharing the data gathered in clinical trials honors the sacrifice made by the trial participants, without whom there would be

The New England Journal of Medicine
Downloaded from nejm.org on May 17, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

no data to share. We appreciate the concern of some that reidentification of participants could be possible, but the funders of clinical research do not see this as an insurmountable barrier, and neither do we. The issue merits reasoned debate. We are happy that the community has joined in, but for the debate to be productive, it should be carried out in the spirit of improving human health; that needs to remain our foremost challenge.

Dan L. Longo, M.D.
Jeffrey M. Drazen, M.D.

Since publication of their article, the authors report no further potential conflict of interest.

DOI: 10.1056/NEJMc1602586

Challenges for Sharing Data from Embedded Research

TO THE EDITOR: The International Committee of Medical Journal Editors recently proposed that authors should be required to share clinical trial data,[1] and both the National Institutes of Health (NIH) and the Patient-Centered Outcomes Research Institute are advancing data-sharing requirements. We fully support these changes for individually randomized clinical trials when participants authorize data sharing. However, this framework poses challenges for cluster-randomized trials that use routinely collected health care data. Such trials are an important focus of the "learning health system." For example, 8 of 10 demonstration projects sponsored by the NIH Health Care Systems Research Collaboratory involve cluster randomization.

Cluster-randomized trials and other types of embedded research differ from conventional trials in three new features: health systems and providers become incidental research subjects, trials can use every diagnosis, procedure, and medication to perform risk adjustment, and trials can involve hundreds of thousands of people who do not give explicit consent. Thus, the potential for unauthorized reidentification is higher,[2] and the consequences may be more severe. It is often impossible to "deidentify" provider organizations, which may have legitimate concerns about the unintended consequences of the disclosure of such information (e.g., adverse marketing by pharmaceutical manufacturers). This consideration may dissuade clinicians and health systems from participating. In addition, the increasing use of distributed analysis methods, in which investigators never have possession of individual-level data, means that there is no conventional data set to share.[3]

An approach to addressing these issues may be to create data enclaves that allow investigators to conduct analyses without taking possession of data while also ensuring that new analyses are consistent with organizations' participation agreements. This approach is used by the Centers for Medicare and Medicaid Services Virtual Research Data Center. Analyses of distributed data sets will require new technical infrastructure and funding. We are convinced it is possible to achieve meaningful data sharing with embedded research that encourages — rather than discourages — the growth of a learning health system.

Richard Platt, M.D.
Joakim Ramsberg, Ph.D.

Harvard Pilgrim Health Care Institute
Boston, MA
richard.platt@gmail.com

Disclosure forms provided by the authors are available with the full text of this letter at NEJM.org.

This letter was published on April 20, 2016, at NEJM.org.

1. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med 2016;374:384-6.
2. Sweeney L. Simple demographics often identify people uniquely (data privacy working paper 3). Pittsburgh: Carnegie Mellon University, 2000 (http://dataprivacylab.org/projects/identifiability/paper1.pdf).
3. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. Health Aff (Millwood) 2014;33:1178-86.

DOI: 10.1056/NEJMc1602016

The New England Journal of Medicine
Downloaded from nejm.org on May 17, 2016. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.