

EDITORIALS



Data Sharing

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

The aerial view of the concept of data sharing is beautiful. What could be better than having high-quality information carefully reexamined for the possibility that new nuggets of useful data are lying there, previously unseen? The potential for leveraging existing results for even more benefit pays appropriate increased tribute to the patients who put themselves at risk to generate the data. The moral imperative to honor their collective sacrifice is the trump card that takes this trick.

However, many of us who have actually conducted clinical research, managed clinical studies and data collection and analysis, and curated data sets have concerns about the details. The first concern is that someone not involved in the generation and collection of the data may not understand the choices made in defining the parameters. Special problems arise if data are to be combined from independent studies and considered comparable. How heterogeneous were the study populations? Were the eligibility criteria the same? Can it be assumed that the differences in study populations, data collection and analysis, and treatments, both protocol-specified and unspecified, can be ignored?

A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group's data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as “research parasites.”

This issue of the *Journal* offers a product of data sharing that is exactly the opposite. The new investigators arrived on the scene with their own ideas and worked symbiotically, rather than parasitically, with the investigators holding the data, moving the field forward in a way that neither group could have done on its own. In this case, Dalerba and colleagues¹ had a hypothesis that colon cancers arising from more primitive colon epithelial precursors might be more aggressive tumors at greater risk of relapse and might be more likely to benefit from adjuvant treatment. They found a gene whose expression appeared to correlate with the expression of genes that characterize more mature colon cancers on gene-expression arrays and whose product was reliably measurable in resected colon cancer specimens by immunohistochemistry. To assess the clinical value of this potential biomarker, they needed a sufficiently large group of patients whose archived tissues could be used to assess biomarker expression and who had been treated in relatively homogeneous way.

They proposed a collaboration with the National Surgical Adjuvant Breast and Bowel Project (NSABP) cooperative group, a research consortium funded by the National Cancer Institute that has conducted seminal research in the treatment of breast and bowel cancer for the past 50 years. The NSABP provided access to tissue and to clinical trial results on an individual patient basis. This symbiotic collaboration found that a small proportion (4%) of colon cancers did not express the biomarker and that the survival of patients with those tumors was poorer than that of patients whose tumors expressed the biomarker. Furthermore, when the effect of adjuvant chemotherapy was assessed, nearly all

the benefit from adjuvant treatment was within the biomarker-negative group, the patients with the most primitive tumors. The findings have generated a new hypothesis that is now ready for testing in a prospective randomized clinical trial.

If the hypothesis that nearly all the benefit from adjuvant chemotherapy is in the biomarker-negative group is confirmed, over 90% of patients with stage II colon cancer will be reassured that avoiding the unpleasantness of standard adjuvant therapy is unlikely to affect their outcome adversely. No one expected that.

How would data sharing work best? We think it should happen symbiotically, not parasitically. Start with a novel idea, one that is not an obvious extension of the reported work. Second,

identify potential collaborators whose collected data may be useful in assessing the hypothesis and propose a collaboration. Third, work together to test the new hypothesis. Fourth, report the new findings with relevant coauthorship to acknowledge both the group that proposed the new idea and the investigative group that accrued the data that allowed it to be tested. What is learned may be beautiful even when seen from close up.

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

1. Dalerba P, Sahoo D, Paik S, et al. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *N Engl J Med* 2016; 374:211-22.

DOI: 10.1056/NEJMe1516564

Copyright © 2016 Massachusetts Medical Society.

Prognostic Subgroups among Patients with Stage II Colon Cancer

C. Richard Boland, M.D., and Ajay Goel, Ph.D.

Colon cancer has traditionally been treated surgically. However, many cases of colon cancer are systemic at the time of diagnosis, and apparently curative surgery is followed at a later date by tumor recurrence as a consequence of circulating tumor cells before the surgery. Adjuvant medical therapies are designed to prevent recurrences after surgical resection.

The current standard for clinical prognostication relies principally on pathological staging. In early-stage colon cancers (stages I and II), all the tumor is contained within the wall of the colon. Less than 10% of patients with stage I disease have a recurrence, and adjuvant chemotherapy is not administered because it provides no benefit. Approximately 20% of patients with stage II colon cancers have a recurrence, and adjuvant chemotherapy provides a minimal benefit¹ that is usually considered to be not worth the toxic effects of the drugs. Furthermore, among patients with stage II disease, there is quite a range of rates of 5-year survival (as low as 60%), depending on the incremental depth of tumor invasion.² Stage III colon cancers have regional lymph-node metastases; cancer recurs in more than 50% of patients, and multiple clinical studies have shown significant increases in survival with the administration of adjuvant chemotherapy.² Patients

with distant metastases (stage IV colon cancer) receive a variety of more intensive regimens that are not usually curative. Rectal tumors are different and are not part of this discussion.

The use of adjuvant chemotherapy in patients with stage II colon cancer remains controversial because recurrence never develops in the vast majority of these patients.¹ Treatment of all patients with stage II colon cancer is “overtreatment,”³ since only a small subgroup of patients derives any therapeutic benefit, whereas in others there is harm, a poorer quality of life, and no net benefit.⁴ Prior studies to identify the subgroup of patients with high-risk stage II colon cancer have not been robust,⁵ and the lack of prognostic and predictive criteria underscores the need to discover biomarkers that can facilitate the selection of patients for additional treatment.

In this issue of the *Journal*, Dalerba and colleagues⁶ report on a novel approach to the problem of identifying patients with colon cancer who might benefit from adjuvant chemotherapy. They reasoned that the presence of a stem-cell-like state would be associated with more aggressive tumors, and they performed a bioinformatics search for a gene-expression signature obtained from populations of stem cells and progenitor cells. By mining a large preexisting