# Missing Visits, Missing Benefits

**Dale Lehman, Loras College**

***Introduction***

The main part of this paper consists of my submission to the SPRINT competition held by the New England Journal of Medicine (NEJM) in February, 2017 (the competition is described in Burns and Miller, 2017). The competition was an experiment in the public sharing of data from a clinical trial – the SPRINT trial (The SPRINT Research Group, 2017). That trial received a great deal of publicity, as it involved a common and serious health condition – high blood pressure. Traditionally, a systolic blood pressure (SBP) of 140 has been the target for controlling high blood pressure. Some health experts have promoted a more aggressive target of 120 SBP. The SPRINT trial was aimed at examining the benefits and risks of using this more aggressive SBP target. The principle results were that the risk of primary events (primarily heart attacks) were significantly reduced, and the risk of related serious adverse events (related to the additional medications needed to more aggressively control blood pressure) significantly increased. This clinical trial received a great deal of publicity, as well as a fair amount of criticism (mostly concerning the tradeoff between benefits and risks, which many saw as not favorable enough to recommend more intensive treatment).

The competition had two rounds: a Qualifying Round which asked participants to choose one of two factual questions to answer in order to qualify for the second round, the Challenge Round. 279 teams started the competition (by requesting the data, which also involved getting IRB approval) and 143 teams completed. Three winners were chosen (*Disclosure: I was not one of them*), and they presented their work at a conference sponsored by the NEJM in May, 2017 where the general issue of public disclosure of clinical trial data was the focus.

This paper is organized as follows: the next section contains a Lay Summary and expanded Abstract which comprised the main submission in the Challenge Round. This is followed by a section describing the Qualifying round. The final section contains my thoughts about the competition and subsequent conference, including the ways that I utilized JMP in my analysis.

## Challenge Round Submission

*Lay Summary*

Nearly 25% of the SPRINT participants missed some scheduled visits, and we find that this group showed no significant benefit from intensive treatment. This group also had increased risk of serious adverse events. Conversely, patients who attended all scheduled visits had even better outcomes than the published effectiveness of intensive treatment in SPRINT, but had increased risk of serious adverse events. Further study may determine interventions that can improve outcomes for the missed visit group.

*Abstract*

To assess the impact of missed visits, we divided SPRINT participants into 3 groups: "early missed" participants missed at least one scheduled visit in the first 3 months (9.6%), "later missed visit" participants who only missed visits after the first 3 months (15.3%), and "no missed visit" participants who made all scheduled visits (75.1%).

The table below reports the relevant group comparisons in terms of the risk ratios for primary events (6% of all participants) and related serious adverse effects (3.6% of all participants):

| Group Comparisons | | Risk Ratio: Primary Event (95% CI) | Risk Ratio: Serious Adverse Effect (95% CI) |
|---|---|---|---|
| **Original SPRINT Results With and Without Additional Factors** | Intensive/Control in original SPRINT publication | 0.75* (0.64, 0.89) | 1.88* (1.51, 2.36) |
| | Intensive/Control with additional factors[1] | 0.74* (0.63, 0.88) | 1.87* (1.50, 2.35) |
| **Treatment Response with Early Missed Visits** | Intensive with early missed visit/ Intensive with no missed visit | 2.14* (1.46, 3.13) | 1.17 (0.72, 1.92) |
| | Intensive with early missed visit/ Control with no missed visit | 1.50* (1.04, 2.18) | 2.35* (1.41, 3.93) |
| **Head-head Comparisons** | Intensive/Control – no missed visits in both | 0.70* (0.57, .087) | 2.00* (1.52, 2.63) |
| | Intensive/Control – early missed visits in both | 1.01 (0.62, 1.64) | 1.05 (0.54, 2.01) |
| | Intensive/Control – later missed visits in both | 0.79 (0.56, 1.12) | 2.11* (1.30, 3.45) |

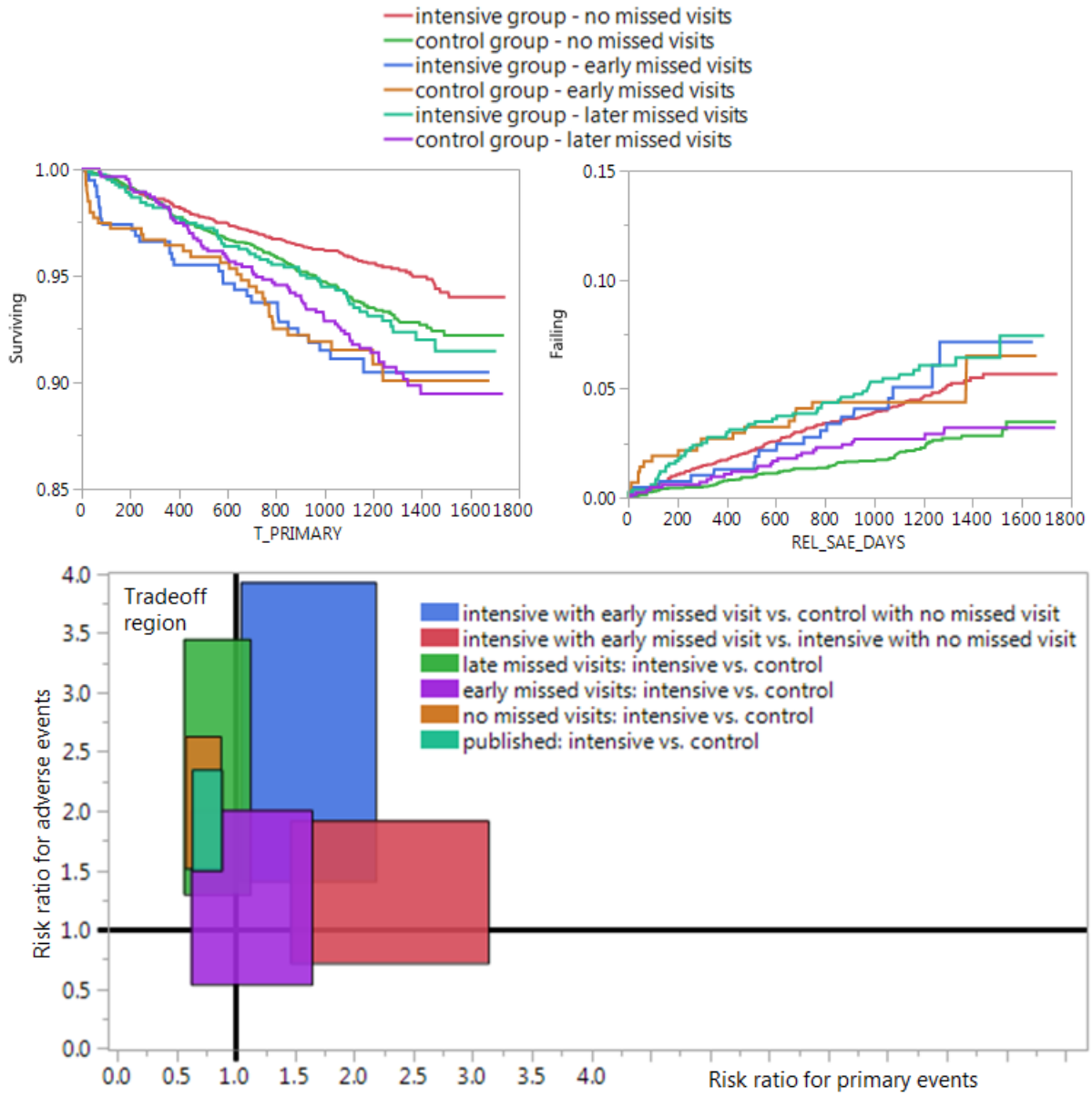
*indicates significant p<.05.

The robust response to intensive treatment in SPRINT disappears for participants that miss scheduled visits within the first 3 months. Missing early visits increases the risk of primary events for both intensive and control groups and is associated with increased risk of serious adverse events related to treatment for both groups.

The group comparisons and risk tradeoffs can be seen in Figure 1, which shows the survival curve for primary events, the failure curve for serious adverse events, and a risk tradeoff with confidence regions:

[1] Including baseline SBP, NOAGENTS, Smoking, EFGR, SCREAT, Race, Age, Female, CHR, and HDL in the Proportional Hazards Model were mostly significant and of the expected sign. Replacing this list with the Framingham Risk Scores was just as effective, so the reported analysis in the last five rows use FRS.

Figure 1: Survival, Failure, and Risk Tradeoffs[2]




In general, the effectiveness of intensive treatment – with no missed visits – is evident, but so is the lack of effectiveness with missed visits. Early missed visits are associated with higher failure rates (serious adverse events), particularly for intensive treatment.

The bottom panel of Figure 1 illustrates both uncertainty in the estimates and the tradeoff between effectiveness and risk of adverse events. Vertical and horizontal lines are drawn at risk ratios = 1. The top right quadrant shows higher risks of both primary and adverse events – an unfavorable outcome. Conversely, the lower left quadrant represents positive outcomes in both dimensions. The top left

[2] Note: the bottom of figure 1 was created using the custom map feature in JMP.

quadrant shows the tradeoff between intensive treatment effectiveness and the risk of adverse events. In that region the darker green rectangle shows a confidence region for this tradeoff using the original SPRINT results. The sides of the rectangles represent 95% confidence intervals for each risk ratio, so if the two are independent the rectangle represents a 90.25% confidence region. The gold region shows intensive treatment compared to control, with no missed visits.

The large uncertainty associated with both risk ratios is evident for groups with missed visits (due to smaller sample sizes), and so is the relative inconclusiveness of intensive treatment effectiveness. In particular, the blue rectangle compares intensive treatment with early missed visits to the control group with no missed visits – here intensive treatment appears to result in unfavorable outcomes.

Patient compliance is known to be a serious impediment to effective treatment.[3] Estimates of noncompliance are on the order of 50%, and 25% of participants with missed scheduled visits in SPRINT are consistent with expectations that clinical trials experience better compliance than real world practice. Early missed visits provide an early warning signal that intensive treatment may not be effective. We found little relationship between missing early visits and baseline factors (other than race). Further study is warranted concerning what early missed visits proxy for, and what interventions might improve outcomes for this group.

Without further data, it is not possible to determine what the missed visits represent. It is plausible that participants who miss visits are not following their provider's instructions. If that is the case, then it is an early warning sign of a patient who may not benefit from more aggressive treatment. On the other hand, they may benefit from additional interventions aimed at increasing compliance. On the other hand, patients may miss visits because they don't feel well – possibly due to the side effects of their medications. Then, the relatively bad outcomes of aggressive treatment are more likely to be biological or chemical rather than behavioral.

Presumably the SPRINT protocol called for additional visits for those whose blood pressure was not under control. Not only was data on these additional visits not provided (not available?), but the missed scheduled visits were never mentioned. Given the sizable segment who miss visits, and the poor outcomes experienced by these patients, this is an area ripe for further investigation. One encouraging finding: more aggressive control of blood pressure is even more beneficial than originally reported – if patients comply with their treatment plans.

**Qualifying Round Submission**

Two questions were provided and entrants could choose either one. The first asked for the test statistic in the Cox Proportional Hazards model. The second asked for the average SBP reading at the last visit of participants in the control and intensive groups. I was leery of the test statistic questions because I believe different software packages often calculate test statistics in slightly different ways. The second question seemed more unambiguous – although it turned out not to be the case.

[3] For example, see Rosenbaum and Shank, 2013, or Osterberg and Blaschke, 2005.

There is some ambiguity in the question since some participants have a last post-baseline visit that is missing an SBP reading. I will provide answers based on these missing values, as well as using the last post-baseline reading that is available for those participants.

The 9361 participants include 112 for whom the baseline SBP is the last reading. That leaves 9249 potential sample size for this question. If I treat the 9 missing last visit SBP readings as missing, there are 9240 total people for whom we have a last post-baseline SBP reading.

| | | |
|---|---|---|
| Control Group | n=4612 | mean SBP = 133.9050 |
| Intensive Group | n=4628 | mean SBP = 119.8835 |

If I use the last available recorded post-baseline SBP reading,

| | | |
|---|---|---|
| Control Group | n=4616 | mean SBP = 133.9053 |
| Intensive Group | n=4632 | mean SBP = 119.8895 |

(note: the total sample size only increases by 8 people rather than 9 in the second calculation due to the fact that one participant had a recorded visit with a missing SBP but had no recorded post-baseline reading, so they are omitted in either case).

In the process of developing my answer, I first attached a Value Ordering to the visit codes (which used RZ for the baseline visit, and then 1M, 2M, 3M, 6M, etc. for subsequent visits at monthly intervals). Since we were interested in the last visit, I ordered them in reverse chronological order. Figure 2 shows the first to participants data:

Figure 2: Excerpt of Blood Pressure Readings

| | | Group | |
|---|---|---|---|
| | | Control | Intensive |
| | | SBP | SBP |
| MASKID | VISITCODE | Sum | Sum |
| S00007 | 51M | 129 | . |
| | 48M | 154 | . |
| | 45M | 128 | . |
| | 42M | 125 | . |
| | 39M | 137 | . |
| | 36M | 144 | . |
| | 33M | 143 | . |
| | 30M | 121 | . |
| | 27M | 117 | . |
| | 24M | 128 | . |
| | 21M | 113 | . |
| | 18M | 120 | . |
| | 15M | 167 | . |
| | 12M | 144 | . |
| | 9M | 144 | . |
| | 6M | 145 | . |
| | 3M | 136 | . |
| | 2M | 117 | . |
| | 1M | 124 | . |
| | RZ | 145 | . |
| S00010 | 36M | 155 | . |
| | 33M | 130 | . |
| | 30M | 132 | . |
| | 27M | 135 | . |
| | 24M | 146 | . |
| | 21M | 142 | . |
| | 18M | 169 | . |
| | 15M | 140 | . |
| | 12M | 139 | . |
| | 9M | 137 | . |
| | 6M | 147 | . |
| | 3M | 131 | . |
| | 2M | 137 | . |
| | 1M | 154 | . |
| | RZ | 138 | . |

I then joined this data set with the original data (dropping multiples) so that only the last SBP reading would be joined. Alternatively, I could have stripped off the numbers from the month codes and used the max(SBP, by MASKID). I suspect may participants proceeded like that, using basic database commands to extract the latest (maximum month) SBP reading. It was also easy to see, from this layout, that the variability of SBP readings might be of interest. Indeed, it did have a significant relationship with the risk of heart attacks, and this was also noted by almost a third of the entrants in the competition. What nobody else noticed, was the fact that many participants had missed visits during

their participation in the trial. This was obvious when I tabulated the data differently, as shown in Figure 3:

Figure 3: Alternative Tabulation View

| | VISITCODE | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 54M | 51M | 48M | 45M | 42M | 39M | 36M | 33M | 30M | 27M | 24M | 21M | 18M | 15M | 12M | 9M | 6M | 3M | 2M | 1M | RZ |
| | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP | SBP |
| MASKID | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum | Sum |
| S00048 | . | . | . | . | . | . | 125 | 119 | 129 | 127 | 119 | 118 | 132 | 118 | 122 | 128 | 116 | 124 | 116 | 131 | 132 |
| S00050 | . | . | . | . | . | 116 | 131 | 155 | 115 | 125 | 109 | 124 | 110 | 111 | 115 | 117 | 107 | 139 | 125 | 117 | 137 |
| S00087 | . | . | . | . | . | 118 | 118 | 119 | 119 | 106 | 100 | 118 | 117 | 94 | 184 | 134 | 129 | 172 | 159 | 146 | 191 |
| S00092 | . | . | . | . | . | . | . | . | 118 | 139 | 137 | 145 | 124 | 150 | 148 | 145 | 153 | 142 | 146 | 150 | 135 |
| S00103 | . | . | . | . | . | . | . | . | 111 | 131 | 125 | 111 | 108 | 111 | 127 | 120 | 125 | 121 | 120 | 106 | 133 |
| S00109 | . | . | . | . | . | 114 | 114 | 100 | 112 | 144 | 93 | 115 | 96 | 119 | 103 | 113 | 123 | 86 | 103 | 122 | 137 |
| S00122 | . | . | 124 | 129 | 147 | 140 | 128 | 141 | 131 | 125 | 127 | 129 | 125 | 115 | 124 | 134 | 140 | 136 | 137 | 133 | 137 |
| S00128 | . | . | . | . | 133 | 115 | 135 | 130 | 130 | 116 | 136 | 115 | 146 | 126 | 129 | 142 | 141 | 127 | 121 | 140 | 125 |
| S00163 | . | . | . | . | . | . | . | 120 | 105 | 112 | 110 | 110 | 109 | 105 | 160 | 119 | . | 119 | 121 | 127 | 120 |
| S00173 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 134 | 146 | 127 | 133 | 151 |
| S00187 | . | . | . | . | 126 | 129 | 142 | 146 | 122 | 149 | 154 | 129 | 140 | 128 | 131 | 150 | 135 | 129 | 130 | 158 | 149 |
| S00193 | . | . | 104 | 105 | 116 | 112 | 87 | 88 | 95 | 110 | 97 | 92 | 100 | 98 | 110 | 107 | 111 | 131 | 111 | 137 | 147 |
| S001E5 | . | . | 114 | 146 | 141 | 120 | 120 | 131 | 126 | 122 | 127 | 137 | 137 | 118 | 117 | 123 | 119 | 129 | 128 | 119 | 124 |
| S00200 | . | . | . | . | 140 | 134 | 142 | 136 | 139 | 134 | 145 | 129 | 119 | 137 | 130 | 152 | 144 | 120 | 111 | 148 | 152 |
| S00201 | . | . | . | . | . | 136 | 136 | 136 | 135 | 136 | 136 | 136 | 130 | 115 | 115 | 123 | 148 | 139 | 139 | 132 | 143 |
| S00242 | . | . | . | . | 121 | 121 | 106 | 119 | 111 | 115 | 132 | 119 | 119 | 119 | 119 | 119 | 120 | 125 | 132 | 152 | 159 |
| S00248 | . | . | . | . | . | . | . | 97 | 106 | 109 | 109 | 105 | 119 | 129 | 126 | 111 | 129 | 116 | 121 | 133 | 138 |
| S00280 | . | . | . | . | . | 149 | 146 | 151 | 140 | 139 | 141 | 136 | 144 | 132 | 133 | 141 | 130 | 136 | 147 | 147 | 141 |
| S00296 | . | . | . | . | . | . | . | 94 | 116 | 137 | 144 | 103 | 99 | 136 | 108 | 106 | . | 130 | . | 102 | 110 |
| S00304 | . | . | 117 | 104 | 119 | 117 | 115 | 119 | 111 | 118 | 114 | 115 | 105 | 115 | 113 | 109 | 121 | 127 | 136 | 127 | 150 |
| S00337 | . | . | . | . | 105 | 141 | 164 | 145 | 129 | 152 | 141 | 143 | 135 | 130 | 129 | 155 | 148 | 135 | 133 | 132 | 117 |
| S00341 | . | . | . | . | . | . | . | 140 | 144 | 157 | 152 | 135 | 147 | 135 | 143 | 123 | 140 | 154 | 134 | 152 | 166 |
| S00343 | . | . | . | . | 122 | 134 | 120 | 120 | 115 | 103 | 116 | 119 | 117 | 118 | 129 | 117 | 134 | 129 | 133 | 137 | 180 |
| S00349 | . | . | . | . | 137 | 170 | 153 | . | . | 129 | 153 | 139 | . | 150 | 149 | 120 | 133 | 133 | 112 | 143 | 164 |
| S00356 | . | . | . | . | . | . | . | 143 | 128 | 115 | 123 | 126 | 119 | 122 | 118 | 95 | 99 | 123 | 118 | 123 | 130 |
| S00363 | . | . | . | . | . | 139 | 149 | 135 | 141 | 132 | 137 | 132 | 135 | 145 | 136 | 138 | 135 | 132 | 139 | 114 | 134 |
| S00397 | . | . | . | . | 137 | 147 | . | 157 | 138 | 147 | 161 | 154 | 134 | 136 | 160 | 139 | 139 | 122 | 136 | 153 | 134 |
| S00406 | . | . | 127 | 140 | 127 | . | 134 | 142 | 138 | . | 136 | . | 135 | 127 | . | 139 | 143 | 139 | 132 | 126 | 128 |
| S00435 | . | . | . | . | . | . | 113 | 133 | 137 | 159 | 138 | 170 | 141 | 127 | 117 | 136 | 135 | 115 | 126 | 103 | 154 |
| S00458 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| S00470 | . | . | . | . | . | . | . | . | 128 | 136 | 135 | 123 | 134 | 123 | 133 | 119 | 122 | 141 | 114 | 120 | 149 |
| S00516 | . | . | . | . | . | . | . | 112 | 114 | 118 | 106 | 111 | 118 | 141 | 177 | 141 | . | 135 | 148 | 157 | 136 |
| S00524 | . | . | . | . | . | . | . | 108 | 127 | 110 | 115 | 113 | 119 | 119 | 122 | 113 | 115 | 122 | 140 | 124 | 132 |

Clearly, some intermediate visits did not have blood pressure readings. This prompted me to look at the detailed protocol for the trial. Participants who exceeded their target blood pressure were supposed to be seen monthly until their blood pressure was under control. Data for these unscheduled visits was not provided, and the meaning of the missed readings was not clear. I inquired about this and was told that the original study did not use any other information, so it was not available. I was not informed whether the data existed or had not been collected. In view of my findings, the issue of missed visits and unscheduled visits looms large.

It was this view that prompted me to investigate the missed visits and their relationship with the trial outcomes. I looked at how many visits each participants missed, the fraction of visits they missed, and I also focused on early missed visits (in the first 3 months). Since the competition was designed to seek new clinically relevant findings, I thought a focus on early visits was prudent. This is also a reason I downplayed the relevance of the variability of SBP readings for each participant – while it was a significant factor, it is of limited use since it cannot be determined early enough to be of clinical use.[4]

[4] There is also a technical issue with using a measure such as the standard deviation of SBP readings. By definition, the intensive group will have a higher standard deviation than the control group, since their blood pressure is

**Observations about the SPRINT Competition**

The winning entries (the top 2) were focused on communicating the results of SPRINT to clinicians, including an "app" that weighs potential decreased risk of heart attacks against potential increased risk of serious side effects of medications. While these were useful and flashy, they were based on point estimates of the effects. As my analysis shows, there is considerable uncertainty in the effects – focusing on point estimates is not necessarily good medical practice. Clinicians need to make decisions – they don't have the luxury of analysts to just study the issue further. But surely the degree of uncertainty is a relevant clinical factor. None of the other entries appeared to focus on uncertainty.

Opening clinical trial data to public use is an important step. The conference had two panels of SPRINT trial participants – they expressed shock that there was any question about making the data available to others. Only the research community appeared to be hesitant, and most of their concerns were due to professional protection of careers. Collecting good data is hard work – people should be rewarded for creating the data, not just analyzing it. That relatively simple change would remove the disincentive to share, since credit would be earned by creating useful data.

Hesitancy about sharing data was evident throughout the competition, beginning with the NEJM itself. As stated in an Editorial (Longo and Drazen, 2016):

> "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."

While the Journal is to be commended for sponsoring this competition, and has been generally supportive of data sharing, it sometimes seems to be placing roadblocks. This was evident in the SPRINT competition, with the insertion of a Qualifying Round, insistence on IRB approval for participation (even though the data had been anonymized and the results already published), and failure to provide data not used by the original researchers – that seemed a lame excuse for not providing requested data in an exercise to see the potential benefits of data sharing.

The process of the competition also underscored the resistance to data sharing. I repeatedly asked questions to the forum in the competition – but few others did. I was used to such discussion from other data competitions (e.g., Kaggle competitions). But clinical trials in medicine are different – researchers are highly competitive and protective. A lot of grant money and professional stature is at stake. It is unfortunate that this competition gets in the way of open data. I view this as a call to fix the incentive structure in academia generally, and medicine, in particular. Let's give credit for collecting the

---

being reduced by something generally greater than 140 to something less than 120 (rather than the 140 for the control group).

data, make it as open as possible, and give further credit to the creators when their data is used by others. Focusing on publishing the results of analysis leads to many of the ongoing issues with lack of replication and flawed research.[5]

**References**

Nancy S. Burns and Panela W. Miller, June 8, 2017, "Learning What We Didn't Know – The SPRINT Data Analysis Challenge, *New England Journal of Medicine*.

The SPRINT Research Group, 2015, "A randomized trial of intensive versus standard blood-pressure control," *New England Journal of Medicine*.

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D., 2016, "Data Sharing," *New England Journal of Mediciine*.

Rosenbaum and Shank, 2013, "Taking Our Medicine-Improving Adherence in the Accountability Era," *NEJM*, or Osterberg and Blaschke, 2005, "Adherence to Medication," *NEJM*.

Nature, 25 May, 2016, "1500 scientists lift the lid on reproducibility," accessed at http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970.

[5] For example, see Nature, 2016 for some recent survey results. A growing number of publications are describing the extent of the reproducibility "crisis" and investigating what can/should be done about it.