

Data-Driven Monitoring of Clinical Trials in the COVID-19 Era

Michael Proschan, Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases
(joint work with Martha Nason, Ana Ortega-Villa, and Jing Wang, 2022,
and separate work with Dean Follmann, 2022)

Clinical trials are monitored over time for safety, general conduct, efficacy, and futility. It is well understood that repeated examination of efficacy data without adjusting for multiple comparisons inflates the familywise error rate (FWER), the probability of falsely declaring a treatment benefit at some point during the trial. Control of the FWER requires knowledge of the joint distribution of standardized z-statistics $Z(t)$ over time t , and this joint distribution has been shown to be asymptotically the same for continuous, binary, and survival endpoints. The key is to define t in terms of proportion of information. Monitoring boundaries have been developed to control the FWER at the desired level under the assumption that the monitoring schedule is independent of $Z(t)$, $0 \leq t \leq 1$. Changing monitoring times in response to observed data trends can inflate the type 1 error rate. We consider two settings that involve such data-driven monitoring times. The first concerns rapidly evolving information in a pandemic such as COVID-19. Trials initially planned with only a single interim and final analysis were changed in response to unexpectedly high efficacy of mRNA vaccines. We show how to seamlessly change monitoring schedules in response to internal data with no possibility of inflating the type 1 error rate. The second setting is monitoring a secondary endpoint at the time the primary endpoint crosses its monitoring boundary. Spending full level alpha on the secondary endpoint at that time seems to follow a gatekeeping procedure, and gatekeeping procedures control the FWER. Nonetheless, we show that in the most extreme case, the FWER can be inflated to the same degree as from repeated monitoring of a single endpoint with no adjustment for multiple comparisons.