Souvenir T: Even Big Data Calls for Theory and Falsification

> Historically, epidemiology has focused on minimizing Type II error (missing a relationship in the data), often ignoring multiple testing considerations, while traditional statistical study has focused on minimizing Type I error (incorrectly attributing a relationship in data better explained by random chance). When traditional epidemiology met the field of GWAS, a flurry of papers reported findings which eventually became viewed as nonreplicable. (Lambert and Black 2012, p. 199)

This is from Christophe Lambert and Laura Black's important paper "Learning from our GWAS Mistakes: From Experimental Design to Scientific Method"; it directly connects genome-wide association studies (GWAS) to philosophical themes from Meehl, Popper and falsification. In an attempt to staunch the non-replication, they explain, adjusted genome-wide thresholds of significance were required as well as replication in an independent sample (Section 4.6).

However, the intended goal is often thwarted by how this is carried out. "[R]esearchers commonly take forward, say, 20–40 nominally significant signals" that did not meet the stricter significance levels, "then run association tests for those signals in a second study, concluding that all the signals with a p-value ≤.05 have replicated (no Bonferroni adjustment). Frequently 1 or 2 associations replicate – which is also the number expected by random chance" (ibid.). Next these "replicated" cases are combined with the original data "to compute p-values considered genome-wide significant. This method has been

propagated in publications, leading us to wonder if standard practice could become to publish random signals and tell a plausible biological story about the findings" (ibid.).

Instead of being satisfied with a post-data biological story to explain correlations, "[i]f journals were to insist that association studies also suggest possible experiments that could falsify a putative theory of causation based on association, the quality and durability of association studies could increase" (ibid., p. 201). At the very least, the severe tester argues, we should strive to falsify methods of inquiry and analysis. This might at least scotch the tendency Lambert and Black observe, for others to propagate a flawed methodology once seen in respected journals: "[W]ithout a clear falsifiable stance – one that has implications for the theory – associations do not necessarily contribute deeply to science" (ibid., p. 199).