An Indianapolis Colts fan begrudgingly accepts problems with reverse causation.

Data is everywhere and can be comprised of basically everything, but don't tell gamblers or fantasy football enthusiasts. To them, data is only one thing, cash. With the correct information, game predictions are as good as gold. Player and team statistics for the National Football League is massive industry all on its own. Based on the “Data & Information” lecture, football data starts out as semistructured (Onesti). The data is already organized by team, season, date, and player roster prior to being captured as the game is played. It is then added to historical databases with information going back to the start of the league in back in 1920. From there it is accessed by thousands of analysts, consultants, fans, and gambling junkies. Since it is meaningful in a specific way to each user, this processed data would be considered information. There is a distinct problem with this massive amount of information being available to fans. Fans, by definition, have a bias towards their favorite team or player. How useful is the knowledge gained when there is an inherent problem of reverse causation? If the starting point in data analysis is 'my team is the best,' it is easy to work backwards and find supporting data to prove any point in support of 'why we are going to win.'

As a Colts fan who had to live in Patriots country for a few years, I had a personal interest in exposing Tom Brady and the Patriots as cheaters. Sports radio was virtually non-stop ‘Deflategate’ talk this entire offseason. The Colts accused the Patriots of using
under-inflated footballs to gain a competitive advantage in the 2015 AFC Championship game (Hipes). One of the more damning arguments against the Patriots was their exceptionally low fumbling rate. I was quick to jump on board to believe that this data was irrefutable evidence that the Patriots were dirty cheaters. When I ran across Kaiser Fung's article in Harvard Business Review debunking this theory, I read skeptically, thinking that the Boston-based writer had his own bias. Unfortunately, the author's tear-down of professional football data analyst Warren Sharp's argument was spot on.

In the article, Fung takes on the premise that the fumbling statistic correlates to cheating. Fung's basic point is ‘the data say the Patriots are excellent at preventing fumbles. It says nothing about why’ (Fung). The blame for this disconnect is the ‘under-appreciated problem’ of reverse causation. In these problems, the result is known and the analyst works backwards to understand the cause. We learned about outliers in the lecture. New England is certainly, and verifiably, an outlier in that they sit well outside the average on the plays-per-fumbles-lost metric. Fung does not argue this point, but does refute the conclusion that the ability to possess these stats means the Patriots were cheating. Fung states ‘(h)owever suggestive, the data does not prove intent or guilt. It simply describes a statistical phenomenon’ (Fung). As the article continues, more analysis of Sharp's work is done, relating to different variations based on teams that play the majority of their games inside of dome stadiums and the difference between fumble recovery as a metric versus plays per total fumbles. When these key pieces of data are taken into account, the Patriots do not seem to be that much of an outlier. As data becomes more available to the average fan, they are “assigning causes somewhat recklessly, because it makes a good story, or helps confirm our biases” (Fung).
In the end, even Sharp himself had to admit that 'New England's spectacular performance on the metric could be explained in any number of ways, including legitimate ones like perfecting ball security techniques or practicing prevention' (Fung). It is difficult to assess future changes in this bias found in the typical sports fanatic. It is clear that access to data and expert interpretation of that data, will become more and more ubiquitous as cell phones, apps, and the number crunching capabilities of personal computers give the fan more and more power. It will certainly be easier to end arguments and prove points, but I doubt this will ever fully override a passionate fan's reverse causation bias. For example, I know all the statistics in the world would not convince me that the Colts aren't the future of the NFL and their leader Andrew Luck isn't a future Hall of Fame athlete. You might as well start placing your bets for next season's Superbowl now: Colts to win big.
Works Cited

Onesti, Nina. "Data & Information." INFO-I101 Lecture, Bloomington, IN.  
19 October, 2015. Class lecture.

Hipes, Patrick. "NFL To Appeal Deflategate Ruling That Voided Tom Brady's 
nfl-1201513714/>.

Fung, Kaiser. "An Important Data Lesson from an Inconsequential Football Scandal 
https://hbr.org/2015/01/an-important-data-lesson-from-an-inconsequential-
football-scandal/>. 