EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

ECON JOURNAL WATCH 10(1)
January 2013: 40-69

Did Jose Canseco Really Improve the Performance of His Teammates by Spreading Steroids? A Critique of Gould and Kaplan

John Charles Bradbury[1]

LINK TO ABSTRACT

> *I talked about how I taught many of the guys, named and unnamed, everything they needed to know about steroids, and said I shared my knowledge freely as I moved from one team to the next. Whenever anyone wanted to know anything about steroids, he always got the same answer: "Talk to Jose. Jose knows. Jose's your man." So they came, and talked, and asked questions. And I shared everything I knew, with friend and foe alike. …*
>
> *I was like a goodwill ambassador, the Godfather of Steroids, and I was genuinely glad to be of help…. So I spread the wealth. I was happy to do it. I wanted to share and I did so hundreds of times, too many times to count.*
>
> —Jose Canseco, *Vindicated* (2008, 2)

Baseball, like many other sports, lends itself to the study of economic behavior, because of the obvious goals of participants, rule constraints, and availability of productivity measures. Eric Gould and Todd Kaplan (2011) use the performances of teammates of the known steroid user Jose Canseco. Jose Canseco's career spanned many years and many teams. Gould and Kaplan use

1. Kennesaw State University, Kennesaw, GA 30144. The author thanks Brad Humphreys and participants at the 2011 Annual Meeting of the Southern Economic Association for helpful comments and suggestions.

a simple econometric strategy to identify the effect Canseco may have had on his teammates. The strategy looks at performances during and after playing as a teammate of Canseco. If Canseco had a positive (or negative) effect on his teammates, then the impact should be observable in player performances.

After examining the performances of a large sample of players over several years Gould and Kaplan conclude, "the empirical analysis shows that a player's performance significantly increases after playing on the same team with Jose Canseco" (2011, 339). Close inspection of the study reveals several serious problems with the analysis, however—problems that result from a flawed empirical approach as well as from their distorted interpretation of their own regression estimates.

The Gould and Kaplan study is developed as an academic contribution on the topic of peer effects among co-workers. But the study has found wider interest because of its use of forensic statistical analysis to identify hidden behavior. Prior to its publication, the study was profiled in the online magazine *Slate* (Fisman 2010). The *Slate* article's main focus is the novel approach of using statistics to identify users of performance-enhancing drugs; the article's subtitle is "Can Statistics Be Used to Find Juicers?" The article was mentioned on the *Freakonomics* weblog (2010a).[2] Statistical analysis was said to support a popular notion that was widely believed in the mainstream media and forcefully advanced by Canseco: that Canseco spread steroids through baseball.

The issue of whether or not Canseco disseminated steroid knowledge through baseball, and whether or not there is evidence of this transmission in the performance of this teammates, is worthy of investigation because conversations about the prevalence of performance-enhancing drugs are being carried on without much scholarly guidance. Evidence that supports the conclusion that Canseco spread steroids through baseball fits with existing notions about players using ergogenic aids to gain an unfair advantage over non-users. These concerns extend beyond baseball. In recent years, cycling and track and field have been plagued with performance-enhancing drug controversies. The increase in hitting and hitting power in the 1990s and 2000s made many of baseball's top performers suspects without substantial evidence. Even players who have not been linked to performance-enhancing drugs in public investigations have been accused of using performance-enhancing drugs because of their physique and hitting prowess (e.g., Brady Anderson, Jeff Bagwell, and Mike Piazza). Thus, external confirmations of

2. I do not mean to criticize either of these publications for writing about this result, because it would be an interesting and useful finding if the results held up. Also, *Freakonomics* published a follow-up post (2010b) that linked to a critique I had published online (Bradbury 2008) of an earlier version of Gould and Kaplan's study.

long-held suspicions and Canseco's own claims, such as the article by Gould and Kaplan, may be used as evidence of steroid use by baseball players of that era. This is not to say that baseball players did not use performance-enhancing drugs or that Jose Canseco did not share his knowledge with other players. My concern is whether careful analysis of historical performance data using statistical analysis offers evidence that supports these popular suspicions.

In this paper, I examine the weaknesses of Gould and Kaplan (2011), identifying methodological problems with the empirical strategy and demonstrating that even if the estimates are taken at face value, they do not support the conclusion that Canseco left a visible trail of steroid users in his wake. I also present a revised empirical model that corrects for the original study's significant defects. My findings are robust: in 77 specifications that I report in this paper, not a single estimate is consistent with the Gould and Kaplan hypothesis, indicating that the authors' reported findings are extremely fragile and thus not indicative of Canseco's presence generating a performance bump. The performance of Canseco's teammates does not indicate a pattern of Canseco spreading performance-enhancing drugs to them.

The Gould and Kaplan approach

Jose Canseco as Johnny Appleseed

Jose Canseco was a 15th-round draft choice of the Oakland Athletics in Major League Baseball's 1982 amateur draft. According to his autobiography (Canseco, 2005), in late 1984, after three seasons of relatively mediocre performance in the low-level minor leagues, Canseco decided to use anabolic steroids in an attempt to improve his performance.[3] In 1985, Canseco improved his performance enough to win *Baseball America*'s Minor League Player of the Year Award, and he was promoted to the major-league club at the end of the season. In 1986, he made the All-Star team and won the American League Rookie of the Year Award. In his 17-year career, Canseco would make six All-Star teams, garner four Silver Slugger Awards, win the American League Most Valuable Player Award, and hit 462 home runs. After his playing days were over, Canseco publicly attributed much of his success to the use of anabolic steroids and growth hormone. He also dubbed

3. Poor performance at the lower levels of the minor leagues does not necessarily indicate poor performance is to be expected at the major-league level. Bradbury (2010a) finds standard performance benchmarks do not predict future performance until players reach the High-A level.

himself “the Godfather of Steroids,” because he taught many other players how to use performance-enhancing drugs.

Canseco’s claims and career make him an ideal subject for studying peer effects among co-workers. Canseco openly used anabolic steroids which have known ergogenic effects.[4] He played for seven different teams over a long career, giving him ample opportunity to share his knowledge with many different players. If Canseco was disseminating knowledge of performance-enhancing steroids to his teammates, there should be an observable Johnny Appleseed effect of players who improved after playing with Canseco. Several metrics exist for measuring aspects of player performance that anabolic steroids should enhance, and the one-on-one contests between hitters and pitchers provide a relatively controlled environment where individual player performance on the field is not heavily affected by teammates during the game.[5]

The basic approach for identifying Canseco’s effect on his peers

The main inferences from Gould and Kaplan (2011) are drawn from empirical analysis using multiple-regression estimated models with the following structure:

$$\text{Performance}_{it} = \beta_0 + \beta_1(\text{playing with Canseco})_{it} + \beta_2(\text{after Canseco})_{it} + \mathbf{X}(\text{other controls})_{it} + \varepsilon_{it} \qquad (1)$$

Performance of player i in year t is estimated to be a function of several parameters, with the main variables of interest being whether or not the player was a teammate of Canseco in year t or had been a teammate of Canseco prior to year t. Tracking several players over many years, some with Canseco as a teammate and some without, should reveal predictable changes in performance if Canseco induced any productivity improvements on his teammates. Thus, the estimates of β_1 and β_2 are of prime importance for identifying Canseco’s role in improving performance through sharing his knowledge of steroids with teammates.

Gould and Kaplan’s justification for using the “with Canseco” and “after Canseco” designations is “since even if a player did learn about steroids from

4. There is ample evidence that anabolic steroids improve strength and power (Hartgens and Kuipers 2004). However, the consensus among medical researchers is that growth hormone—though widely abused by athletes as an alleged ergogenic aid—does not improve strength and power (Stacy, Terrell, and Armsey 2004).

5. Bradbury and Drinen (2008) find that individual hitting performance is largely independent of teammate spillovers.

Canseco, we do not know when he learned about it during his time with Canseco, but we can be sure that he already acquired the knowledge after playing with Canseco" (2011, 341-342). Rather than focus on both the "with Canseco" and the "after Canseco" indicators, the authors place almost all their emphasis on the estimates of the after-effect.[6]

Gould and Kaplan code all years in which a player plays on the same team with Canseco as one, and zero otherwise. For the after-effect indicator, it is not until after the player and Canseco play on different teams that the player is coded as one. The method creates the obvious problem of measuring the impact on players who remained teammates with Canseco for multiple years. Many of Canseco's teammates played with him for several seasons, during which time Canseco himself claims to have used performance-enhancing drugs along with his teammates. For example, teammate Mark McGwire, an admitted steroid user and a player named by Canseco as someone he injected with steroids is coded "with Canseco" from 1986 through 1992. He would not be recognized as receiving Canseco's steroid spillover performance effect until after seven years as teammates. Yet, according to the study's estimates, McGwire was not aided during his time as Canseco's teammate; it was only after he and Canseco separated that he benefitted from Canseco's influence—an odd conclusion. Such delayed spillover applies to many of Canseco's teammates.

The "after-Canseco" interpretation is even more convoluted in the authors' method for players who play on the same team as Canseco but on non-consecutive occasions, involving their playing together on different teams. Players are coded as "with Canseco" for the first year playing with Canseco for every year all the way through to the last year the player is on a team with Canseco. That is, it is not until the first season after the final year of playing with Canseco that the player is coded as "after Canseco."

Table 1 summarizes the empirical estimates for the main variables of interest reported in Gould and Kaplan (2011). The table lists the number of occurrences of "statistically significant" (e.g., p-values less than 0.05) estimates on coefficients for playing with and after Canseco by dependent variable. Positive with-Canseco and after-Canseco coefficients would be consistent with Canseco having a positive performance spillover onto his teammates. The with-Canseco coefficient is not positive and significant in any reported specification; it is negative and significant in two specifications. The after-Canseco coefficient is positive and significant in 12 of the 27 specifications (44 percent).

6. The downplaying of the "with Canseco" indicator is seen, for example, in Gould and Kaplan's not even reporting the with-coefficient estimates in Tables 7, 9, and 10.

TABLE 1. Summary of estimates reported in Gould and Kaplan (2011)

	Estimates where p<0.05 (+/−)	
Dependent Variable (Number of specifications)	With Canseco	After Canseco
Home Runs (4)	1 (−)	3 (+)
Strikeouts (2)	1 (−)	2 (+)
Runs Batted In (2)	0	2 (+)
Slugging Average (3)	0	0
Batting Average (3)	0	1 (+)
Intentional Bases on Balls (2)	0	0
Bases on Balls (2)	0	1 (+)
Steals (2)	0	0
Fielding percentage (2)	0	0
Errors (1)	0	0
At-bats (1)	0	1 (+)
Games Played (1)	0	1 (+)
Earned Run Average (1)	0	0
Inning Pitched (1)	0	1 (+)
Total Positive	0	12
Total Negative	2	0

The authors state that in one non-reported specification using a single with-and-after Canseco indicator produced an estimate that having Canseco as a current or past teammate increased home runs by 1.4 per season. As the other estimates show, however, this is being driven by after-effects, and it is not clear that separating from Canseco as a teammate is a good proxy for identifying a peer effect. Because no alternate specification that employs only a "with Canseco" variable is reported, I am left to speculate that there is no pure "with Canseco" effect. This is a result that conflicts with the spillover hypothesis; yet, it is not portrayed as counter-evidence by the authors. Gould and Kaplan posit three reasons why the observed after-effect is stronger than the with-effect:

> [P]layers who learn about steroids from Canseco do not take steroids during the whole time they are playing "with Canseco," but do use them during the entire time that they are former teammates with him. Alternatively, it may take some time for Canseco's positive effect to be realized, or this pattern may be due to the fact that players who play with him spend more of their time as former teammates of Canseco than being current teammates of him. (342)

Though estimates presented in this paper do not provide evidence of a post-Canseco positive spillover, even if the Gould and Kaplan estimates are taken at face value, they could be interpreted as contrary evidence of Canseco spreading steroid knowledge. The lack of performance improvement while playing with Canseco, especially because many players in this cohort played with Canseco for several years—some of whom admitted using steroids during this time—contradicts the steroid spillover hypothesis. Therefore, the interpretation that the results reflect delayed positive spillovers from Canseco's steroid education program is not the only or best inference that can be drawn from the study. The fact that the "with Canseco" indicator is never positive and significant in 27 estimates reported in the Gould and Kaplan paper is strong evidence against the Canseco effect. While empirical estimates are open to many interpretations, I think it is fair to say Gould and Kaplan's estimates do not provide particularly strong support for the existence of a positive Canseco productivity spillover.

Dependent variables

Gould and Kaplan estimate the impact of having Canseco as a teammate on eight hitting measures (home runs, strikeouts, runs batted in (RBI), slugging average, batting average, intentional bases on balls, bases on balls, and at-bats), one base-running metric (steals), two fielding metrics (errors and fielding percentage), and two pitching metrics (earned run average (ERA) and innings pitched). In attempt to focus on the physical strength boost produced by using steroids, the authors state that they believe home runs, slugging average, RBI, strikeouts (power hitters tend to strike out more often than non-power hitters), and bases on balls (pitchers "pitch around" power hitters) are performance measures that should be particularly sensitive to steroid use for power hitters. The authors predict that steroids would affect non-power hitters in other ways, and thus they use other metrics to search for performance effects. They focus on batting average, steals, errors, and fielding percentage to examine position players.[7] They use ERA to examine pitchers. They posit that games played, at-bats, and innings pitched should capture endurance effects for all players.

7. In Section 3 of the Gould and Kaplan paper, however, the authors claim that steroids should not affect steals or fielding: "These results suggest that Canseco had no effect on measures which clearly should not be affected by steroids (fielding percentage and perhaps steals), but did have an effect on an outcome which is very important for these types of players (batting average)" (343).

In particular, Gould and Kaplan place prime importance on the impact of having Canseco as a former teammate on home runs hit in a season. The authors estimate the impact of Canseco on home run totals more than any other measure presented in the paper, and later use it for comparison with the potential impact of other players on their respective teammates' performances. One problem with using total home runs per season is that the data is positively skewed to an extreme degree, which raises the possibility of strong influence by outliers. This potential bias is never addressed in the paper. A bigger problem, however, is that total output is affected by performance opportunities as well as performance level. Thus, changes in total output over time may be the product of managers choosing to play a player more in the future due to aging, health, or other reasons, outside of any influence steroids might have. Gould and Kaplan briefly address this potential bias by including at-bats as an independent variable in one specification. Despite the statistical significance of at-bats, the control variable is excluded in further analysis because the authors believe the number of at-bats a player receives is endogenous, as evidenced by the size and standard error of the after-Canseco coefficient in the single estimate using at-bats as the dependent variable. However, using raw home run totals without controlling for opportunities is not the best solution. A home run rate normalized for opportunities is a far better dependent variable choice for measuring changes in performance over time, as it avoids the endogeneity issue and does not ignore an obviously relevant factor. The other dependent variables represented as season totals also suffer from the lack of normalization.

While the other dependent variables are popular performance metrics among baseball fans, they are odd choices for measuring performance changes among Canseco's peers to identify steroid use. Effects on strikeouts, RBI, intentional bases on balls, and bases on balls are estimated because these measures are "indicative of a higher performing 'power hitter'" (342). There is no need for "indicative" measures when numerous direct and superior metrics of hitting power exist (I describe these metrics in a later section). Observing variables correlated with power, rather than those measuring power directly, is hardly a robustness test: in some cases, it is double-counting using inferior metrics. It is true that home run hitters tend to strike out more and receive more walks than hitters with less power. But bad hitters also strike out more frequently than good hitters; many hitters with low batting averages walk more to compensate for their diminished hitting power; and many batters are intentionally walked for strategic reasons such as base-out situations and pitcher-batter matchups (e.g., batters who routinely bat before the pitcher are normally poor hitters, but they are frequently walked so that the defense can face the weakest hitter in the lineup). RBI is a performance measure that is largely dependent on batting order position and the quality of batters on a player's team, which pollutes its usefulness as a benchmark of individual achievement.

For this reason, RBI has long been regarded as a poor measure of individual performance.[8] All of these metrics are season totals rather than normalized rates, which means that they are heavily influenced by playing time. In any event, the observations of indicative-of-power variables offer little support for the positive peer effect hypothesis.

Only two offensive dependent variables included in Gould and Kaplan (2011) are not season totals: batting average and slugging average. Batting average is the total number of hits per at-bat. In one of three estimates reported in Gould and Kaplan (2011), the "after Canseco" indicator is positive and statistically significant at better than the standard five-percent level; it is for non-power hitters only. Though batting average is a rate stat, it does not differentiate power from non-power hitting—all hits are weighted the same. The slugging average is a weighted batting average that weights each hit by the number of bases the hitter advances; thus, the slugging average reports the average number of bases per at-bat. The slugging average is never statistically significant in any estimate for Canseco; however, it is positive and statistically significant in estimates of teammate spill-overs for five other comparable hitters reported in Gould and Kaplan (2011): Rafael Palmeiro (Table 6), Ryne Sandberg (Table 6), Matt Williams (Table 9), Chili Davis (Table 9), and Dante Bichette (Table 9). Based on slugging average, several other players are more strongly associated with teammate improvement spillovers than Jose Canseco. This highlights this importance of dependent variable choice in framing the results as supporting or rejecting the Canseco peer effects hypothesis. A thorough discussion of dependent variable choices is presented below.

Control variables

The control variables employed by Gould and Kaplan (2011) are also problematic, and thus call into question the specification choices and the significance of the estimates they produce. If the regression strategy is misspecified, then estimated coefficients that meet the standard threshold of statistical significance do not support hypothesis rejection. Gould and Kaplan do not report estimated coefficients for any variables other than the "with Canseco" and "after Canseco" indicators; therefore, it is unclear what effect and importance the included variables have. I comment briefly on potential problems on included variables in order to explain why specification choices presented in this paper differ from those in Gould and Kaplan (2011), not because I believe the included variables are necessarily contributing to significant bias in the regression estimates. The other

8. See Bradbury (2008) for a discussion of the usefulness of common baseball metrics for measuring performance.

included variables are: the slugging average of division, manager's lifetime winning percentage, ballpark factor, experience, experience squared, and league-year fixed effects.

The offensive output of the division is an odd choice. Major League Baseball divisions are largely an arbitrary construct designed to promote rivalries. Major League Baseball teams have always played games outside their divisions, and after 1996 teams played games outside their leagues. While scheduling does concentrate play within each division, the concentration has changed much over time with the addition of new divisions, interleague play, and unbalanced schedules (all teams within a division do not play identical schedules); therefore, a division's offensive output provides little useful information. Instead, this control variable serves as a poor proxy for the offensive environment of the league, which is partially captured by the league-year fixed effects.

Career wins by a manager provide little information regarding managerial ability, because managerial records are determined heavily by the quality of the team managed. Bradbury (2010b) examines managerial influence on player performance using 30 years of player performances and finds no evidence of managerial effects on player performance.

Using experience instead of age to control for the natural rise and decline of player performance over a career is also problematic. The rise and decline in performance among athletes is physiological, as all athletes exhibit a tendency to improve until peaking in their early twenties to early thirties and then declining (Schulz and Curnow 1988)—a pattern that Bradbury (2009) finds among baseball players. Although age and experience are correlated ($r = 0.89$ among non-pitchers in the sample), there is a bias as to how they covary. Superior baseball players tend to enter the league at younger ages than inferior players; thus, experience is correlated with ability. In addition, stellar players tend to have more experience than less-skilled players at the same age, and more-skilled players exit the league later than inferior players. Such effects in entry and exit mean that using experience as a controlling factor for the rise and decline in performance has the potential to underestimate the impact of aging, because more experienced players tend to be better players and thus remain in the league to exhibit a performance decline.[9]

League-year fixed effects offer some controlling influence for the change in the offensive environment of each league, but they also reflect other factors not related to changing offensive environments altering player performance. For

9. Another problem with using experience to control for aging is that labor rules often affect when a baseball player enters the league. The example of two contemporary Hall-of-Fame third basemen, Wade Boggs and George Brett, demonstrates the issue. Brett entered the league at age 20, while Boggs did not start play until he was 24. At times when their physiological aging functions were the same, Brett is considered four years Boggs' senior.

example, in two seasons during the timeframe of the study, a significant number of games were lost to a labor strike. In 1994, teams played an average of 48 fewer games and in 1995 teams played an average of 18 fewer games. When count totals are used as dependent variables, the lower numbers were a product of playing fewer games with an uncertain effect as to how the propensity for home runs and runs scoring were affected.[10] In rate terms, home runs and runs per game were at historically high levels; however, the raw totals of players will indicate that home run performance declined. With totals as a dependent variable, an explicit control for the offensive performance in each season is needed to be sure that the deviation in performance is offensive-environment related rather than the result of an unrelated exogenous shock like the 1994-1995 strike. Therefore, it is important to control for changes in league performance explicitly, to ensure that such problems are not biasing the main coefficients of interest.

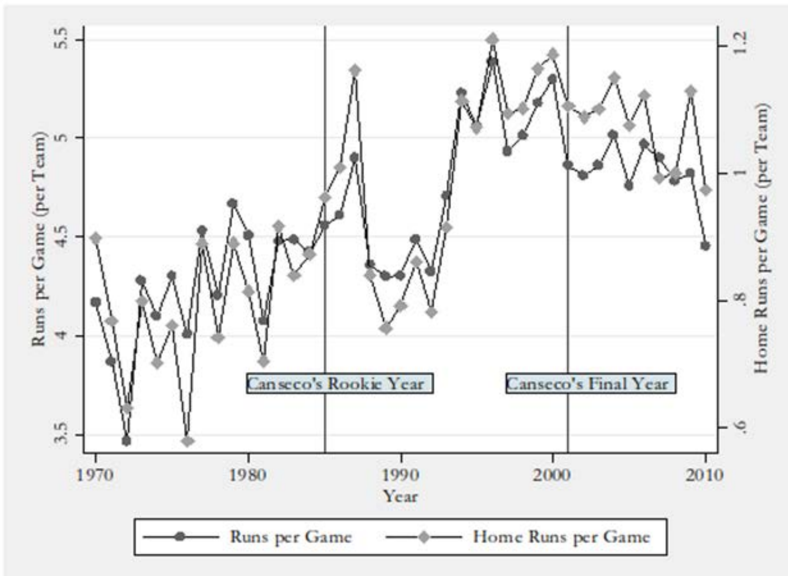
Sample years

Gould and Kaplan use a sample of individual player performances from 1970 to 2009. Canseco's career spanned from 1985 to 2001, which means the sample includes 15 years of player performances before Canseco played and only eight years of performances after his career ended. It is unclear why the 1970 beginning date for the sample was chosen.

The start date of the sample is somewhat problematic because baseball underwent a transition in offense during this period. Figure 1 maps runs per game per team and home runs per game per team from 1970 to 2010 in the American League, where Canseco played his entire career. The league experienced a trend of rising run scoring during the 1970s, due in large part to the introduction of the designated hitter in 1973. After a moderate decline in run scoring in the late 1980s and early 1990s, in the mid-1990s the league experienced a dramatic rise in offense, especially in regard to home runs.

10. Performance rates in these seasons could be affected by less fatigue and lack of spring training preparation, but I suspect the effect would be minor.

Figure 1. American League offense, 1970-2010



Offense tended to be rising for most of Canseco's career; therefore, it is natural to expect anyone who played with Canseco to see his offense improve after leaving Canseco. The authors argue—possibly in anticipation of a critique like the present one—that if such effects are driving the results, then there would be similar effects observed from a sample of ten comparable power hitters, and they describe the estimates for the comparison cohort to be "strikingly different." According to their estimates, however, five other players do have positive after-effects, and I argue below that the estimates are not particularly robust for Canseco. For most estimates, no performance spillovers are identified for Canseco or any other player; thus, I disagree that the estimates are strikingly different. Even so, no justification is offered for extending the sample back this far into the past, and doing so only risks weighting the Canseco coefficients positively by including multiple observations from players many years before Canseco entered the league that adds numerous observations of "0" for the with- and after-effect variables from a low-offense era.

Cutoffs for inclusion in individual years

The cutoff choices of 50 plate appearances for hitters and ten games for pitchers do not make much sense. For hitters, a part-time player who receives only 50 plate appearances will play briefly and face different circumstances compared to full-time players, who typically accumulate 500 to 700 plate appearances in a season. Over the course of a season, a part-time player may play only when matchups are favorable or when the score differential is high. These few observations mean that performance metrics are more likely to deviate from true ability than performance measured over the course of a long season, where fluctuations in performance have more opportunity to even out. In a footnote (342), Gould and Kaplan report that an alternate estimate using a 200 at-bat cutoff produces a similar after-Canseco coefficient estimate. The authors state that they prefer the 50 at-bat cutoff because they believe at-bats are endogenous to playing with Canseco; however, as I state above, normalizing is a superior strategy for handling endogeneity.

For pitchers, the ten-game cutoff means that relievers, who pitch to a small number of batters per game, are included in the sample despite having few opportunities to perform. This is a curious choice considering that in-game opportunities, not games played, are used for hitters. In any event, I do not address pitchers in this paper because pitchers are not the main focus in Gould and Kaplan (2011), and Roger Tobin (2008) finds the impact of steroids is likely to be far less for pitchers than hitters based on the physics of the game.

Arbitrary partitioning of the sample

Gould and Kaplan separate hitters into two classes, "power hitters" and "position players," and they focus on the former group as more likely to be aided by steroid use.[11] The authors define power hitters as players who played the majority of their careers at first base, catcher, any outfield position, or designated hitter. All other players who played a majority of their career games at second base, third base, and shortstop are designated to be position players. My first objection to this partitioning is that it is not necessary, because marginal improvements to hitting will benefit all players no matter what their fielding position is. Many players who are renowned for their hitting power play at "position player" positions, including Alex Rodriguez and Miguel Tejada—both of whom played the majority of their careers at shortstop, were accused by Canseco of using performance-enhancing drugs, and admitted using performance-enhancing drugs.

The unnecessary separation of players into classes becomes more troublesome because the assigned designation of "power" and "position" players is not consistent with the traditional understanding of hitting prowess by position; thus, the reported results are derived from an arbitrary sample and not produced

11. Readers familiar with traditional baseball terminology may be confused by the use of the term "position players," which typically refers to all non-pitchers; however, for the purpose of this paper I will use the term consistent with the designation used by Gould and Kaplan (2011).

by the theoretically most-relevant sample. Fielding ability becomes more important with the frequency with which players must handle fielding chances. At positions where fielding is more important, managers are more willing to trade off offense for defense. In an extensive analysis of the productivity of baseball hitters over baseball history, Michael Schell (2005) describes this relationship as a law of baseball—"The offensive ability at a given position is inversely related to the defensive demands of the position" (199)—and identifies shortstop, catcher, second base, and center field as the positions with the weakest hitting positions for most of the years coinciding with Canseco's career. Bradbury (2010a) reports the most important fielding positions based on fielding chances to be (ordered from most to least important): catcher, shortstop, center field, second base, third base, right field, left field, and first base. This ordering is consistent with the ordering of famed baseball analyst Bill James (1982), who organized positions on the "defensive spectrum" according to positional difficulty. Thus, catcher and center field, two positions that are on the difficult side of the defensive spectrum, and from which hitting expectations are less than other positions, are identified by Gould and Kaplan as power positions, while third base, a position that is (slightly) on the less difficult side of the defensive spectrum, is identified by Gould and Kaplan as a defensive position.

The way the sample is split does not conform to the intention of focusing on a cohort who would most benefit from steroid use because it includes non-power players. It is unclear how the supposed stronger effect among the so-called "power" cohort should be interpreted, because the group designations are not properly composed to meet their intended purpose. Given that this partitioning of the sample in a non-intuitive way produces what the authors feel is favorable evidence of a Canseco peer effect, the sample parsing deserves an explanation and ought to be compared to other arbitrary sample divisions.

Re-estimating Jose Canseco's effect on his peers

The analysis above points out substantial flaws in the method and interpretation of the results of Gould and Kaplan (2011). It is possible that a better empirical model might yield results that support the Canseco hypothesis. Therefore, I use the basic framework employed by Gould and Kaplan, observing player performances over time and correcting for the deficiencies identified above, to identify any spillover effects that Canseco may have had on his teammates.

Empirical method

Data is from Baseball-Databank.org, an open-access database of baseball data.[12] I use the same basic regression function employed by Gould and Kaplan (2011), represented by Equation 1, to estimate the impact of playing with Jose Canseco on performance. However, I use different specifications to identify with- and after-effects of playing with Canseco, and I detail the differences below.

In regard to the dependent Performance variable, I estimate specifications using seven measures of hitting production: home runs per at-bat, home runs per hit, extra-base hits per at-bat, extra-base hits per hit, isolated power, slugging average, and batting average. Though there are many areas of baseball performance, the focus of this analysis is hitting power, because it is the area most likely affected by anabolic steroid use. Hartgens and Harm Kuipers (2004) survey the scientific literature and find short-term administration of anabolic steroids is associated with improvements from baseline strength between five and 20 percent. In an analysis of the impact of strength gains within the range identified by the scientific literature on home run hitting, Tobin (2008) reports, "Basic mechanical principles, in combination with simple but plausible models, show that relatively modest increases in muscle mass, well within the range that can reasonably be expected from steroid use, can dramatically increase home run production" (19-20).

Home runs are the most powerful hits that batters can produce. Batters who hit with power also tend to advance more bases on non-home run hits than do batters who lack power; thus, extra-base hit propensity takes into account extra hitting power that might be gained but is not captured by home runs. Home runs and extra-base hits are normalized per at-bat and per hit. The former measure reports the overall rate of output when a player may or may not hit the ball, while the latter measures the power per hit ball relative to the number of hits. Arthur De Vany (2011) describes the per-hit normalization as the purest measure of hitting power because it excludes other factors that may contribute to the frequency of hitting power outcomes.

Slugging average and batting average are the only dependent variables employed both in Gould and Kaplan (2011) and in this study. Slugging average was originally designed to measure power, but it also includes non-power performance. Power hitters do tend to have higher slugging averages; however, because it is a batting average, high batting average hitters also tend to have high slugging averages. An alternate way to measure hitting power is to subtract the batting average from slugging average to create a metric known as "isolated power."

12. The data and Stata do-files are available on the *Econ Journal Watch* website (link to 9MB zip file).

Isolated power is weighted average of extra-base hits per at-bat. Batting average records the frequency with which a batter hits his way on to base without regard to hitting power, and I include it to capture any non-power effects.

TABLE 2. Summary statistics of variables in regression estimates

Variable	Mean	Standard Deviation	Minimum	Maximum
Home Runs per At-Bat (HR/AB)	0.025	0.019	0	0.153
Home Runs per Hit (HR/H)	0.095	0.073	0	0.522
Extra-Base Hits per At-Bat (XB/AB)	0.077	0.029	0	0.246
Extra-Base Hits per Hit (XB/H)	0.3	0.101	0	0.778
Isolated Power (ISO)	0.132	0.064	0	0.537
Slugging Average (SLG)	0.389	0.088	0.067	0.91
Batting Average (AVG)	0.257	0.04	0.064	0.468
With Canseco	0.015	0.123	0	1
After Canseco	0.048	0.215	0	1
Ballpark Factor	100.2	4.57	88	129
League Runs per Game	4.49	0.38	3.47	5.39
Age	28.78	4.15	18.79	47.85
Age^2	845.27	249.09	353.06	2290.09
Career HR/AB	0.025	0.016	0	0.094
Career HR/H	0.097	0.06	0	0.391
Career XB/AB	0.079	0.022	0	0.148
Career XB/H	0.302	0.076	0	0.65
Career ISO	0.135	0.051	0	0.325
Career SLG	0.395	0.065	0.097	0.624
Career AVG	0.259	0.025	0.083	0.364

The main variables of interest differ from what Gould and Kaplan used to proxy Jose Canseco's impact on his teammates. The "with Canseco" indicator is equal to one during seasons in which the player is a teammate with Canseco and zero otherwise.[13] The "after Canseco" indicator is coded as one in all years *after the first season* as Canseco's teammate and zero otherwise. This variable captures the lagged effect from any learned steroid knowledge by identifying any former *or present* teammate as having at least one year to apply the acquired spillover knowledge. This corrects a major defect in the Gould and Kaplan indicator, which does not identify any after-effect until after leaving Canseco's team for the last time.

13. In 2000, Canseco played a brief stint on the New York Yankees, appearing in 37 games as a part-time player at the end of the season. Canseco writes, "I spent only a couple of months playing for the Yankees, and that's not much time to get to know someone" (Canseco 2005, 230). Therefore, his teammates during this time are not coded as "with Canseco." Gould and Kaplan (2011, 341 n. 11) also do not code his Yankee teammates as "with Canseco."

To control for the talent of each player, I employ either of two strategies. First, player *i*'s career performance in the dependent variable is included as an independent variable to anchor the performance to the player's ability. Second, player fixed effects are employed. However, because fixed effects pick up other potential outside factors—such as the fact that the player was a teammate of Jose Canseco—this may not be the ideal strategy. Both strategies are employed to ensure robustness, and neither the random- nor the fixed-effects estimates identify a positive Canseco effect. Also, because player performance tends to rise and decline with age as a quadratic function, age and age squared are included.

To control for the offensive environment, a ballpark factor for each park and the average runs per game in the league in year *t* are included.[14] Year effects are included in some specifications to control for potential effects unique to the year of observation. Table 2 lists the summary statistics for the included variables.

The models are estimated using two sample time periods. The first time period includes observations from 1970 to 2010, a range that extends the Gould and Kaplan (2011) sample by one year. The second time period ranges from 1982 to 2004, and this range was chosen for two reasons. First, the sample spans four years before and after Jose Canseco's baseball career, so as to cluster observations symmetrically around the time Canseco would spread his steroid knowledge. Second, this sample ends before Major League Baseball instituted a drug-testing program to serve as a deterrent. In 2005, Major League Baseball instituted its first mandatory drug testing system with suspensions for a first offense.

Gould and Kaplan postulate that the peer effect from ergogenic drug knowledge may have disappeared with testing, which they state began in 2003. When testing actually began to impact performance-enhancing drug use is debatable, but I believe 2003 does not represent the pivotal season for a change in drug use. As a part of the 2002 collective bargaining agreement, the 2003 drug tests were implemented *for survey purposes only*—the tests were anonymous and there was no sanction for testing positive.[15] In 2004, a player would have to fail at least two drug tests before being subject to suspensions or fines. Testing in 2003 and 2004 did not result in any suspensions, fines, or public announcements of positive tests for any major-league player: hardly a sign that Major League Baseball was

14. Ballpark factors are provided by Baseball-Databank.org. Factors are calculated based on comparisons of runs scored and allowed in home and away parks by the home and visiting teams to parks.

15. MLB.com offers a timeline of Major League Baseball's performance-enhancing drug policies (**link**). Gould and Kaplan (2011) cite the drop in positive drug tests in 2004 from 2003 as evidence of the effectiveness in testing. However, an alternate explanation for the decline is that the 2003 positive rates were abnormally high as the result of players voluntarily refusing to take tests in 2003. It was reported that some players discussed refusing to take the tests to generate false positives in order to trigger the implementation of more stringent testing in the future; however, to my knowledge, no player publicly admitted to doing so (see Merkin 2003).

identifying and punishing steroid users in a manner that would deter use. It was not until 2005 that Major League Baseball instituted a revised drug testing program that included a ten-day suspension without pay for a first positive test, and several major-league players received suspensions for positive tests, including former Canseco teammate Rafael Palmeiro. In every year since 2005 Major League Baseball has suspended multiple players as a result of testing positive for performance-enhancing drugs. Therefore, 2005 is likely a better time to mark the first season in which Major League Baseball operated a drug-testing program with a credible threat of punishment.

Also, two sample cutoffs are used: 50 plate appearances, which Gould and Kaplan use, and 200 plate appearances, which provides a larger sample of performance per player, which I argue above is preferable. All non-pitchers who meet the cutoffs and played for a full season on the same team are included in the sample. Players who switch teams during the season are excluded for that season, because moving may impact performance and it is unclear how spillover effects may impact performance when the season is split on multiple teams. Jose Canseco is excluded from the sample.

As one further test of Canseco's influence, I examine the performance of eight teammates with whom Canseco claims to have discussed performance-enhancing drug use: Jason Giambi, Juan Gonzalez, Dave Martinez, Mark McGwire, Magglio Ordonez, Rafael Palmeiro, Ivan Rodriguez, and Miguel Tejada.[16] If Canseco shared his wisdom with these players as he claimed, and the estimation strategy for identifying the peer effect is sound, then the boost in performance should be evident in this sample.

The Wooldridge (2002) test for first-order serial correlation in panel data identifies serial correlation in the data. The equations are estimated with fixed and random effects using the Baltagi and Wu (1999) method to correct for serial correlation. Gould and Kaplan do not report detection of or corrections for serial correlation.

Results

Table 3 provides a summary of the estimates of with- and after-Canseco effects, and Tables 4 through 10 report the regression estimates of the multiple specifications for each dependent variable. The estimates for each dependent variable are presented in separate tables, and the row numbers correspond to

16. Canseco implicates Magglio Ordonez in Canseco (2008). All other players are implicated in Canseco (2005). Canseco also implicates several other non-teammates, but they are excluded because of the study's intention to estimate peer effects of co-workers.

similar specifications across tables. In a few instances, estimates were not possible due to the lack of positive definite matrices, which is why model numbers are missing in some tables. In summary, ten specifications were estimated for each dependent variable, and in none is any "with Canseco" and "after Canseco" coefficient positive and statistically significant. In fact, in several cases one or both are negative and statistically significant.

TABLE 3. Summary of impact of Canseco on performance

	Estimates where $p<0.05$ (+/−)	
Dependent variable (number of specifications)	With	After
Home Runs per At-Bat (10)	6 (−)	2 (−)
Home Runs per Hit (10)	1 (−)	0
Extra-Base Hits per At-Bat (10)	7 (−)	0
Extra-Base Hits per Hit (10)	8 (−)	0
Isolated Power (10)	7 (−)	1 (−)
Slugging Average (10)	4 (−)	0
Batting Average (10)	0	0
Total Positive	0	0
Total Negative	33	3

Table 11 reports the estimates of direct spillover effects onto eight hitters whom Canseco identified as teammates he aided with using steroids. The specification reported for each dependent variable corresponds to Specification 5 in Tables 4 through 10. Again, in no specification is any Canseco indicator positive and statistically significant.

Discussion and conclusion

In the analysis of Jose Canseco's peer effect on his teammates, Gould and Kaplan state, "Overall, the evidence points to the strong contagion effect of improper behavior which can be generated by one worker when the incentives to keep up with fellow workers are very strong" (340). The authors take the next step of drawing several implications for other workplaces:

> Outside the world of sports, similar forces may be at work in terms of accounting practices, unprofessional behavior by lawyers, overly aggressive subprime lending, political corruption, public disclosures, cheating by students, accuracy in journalism, reporting in academic research, etc.

> By demonstrating that unethical practices can spread through a contagion effect, our analysis leads to several potential policy implications. The most obvious policy implication is to increase the punishment on individuals practicing unethical behavior and/or transferring their knowledge of such practices to other workers. In addition, policies could be designed specifically to stop the spread of unethical behavior among workers. In particular, the firm (or trade organization) could reward individuals for reporting unethical practices of other workers. …Another possible way of containing a contagion is the use of group punishment for the actions of individual workers (347).

Close examination reveals that Gould and Kaplan's empirical strategy is flawed. I leave the reader to ponder whether that affects the pertinence of the remarks just quoted about unethical behavior and how to stem it.

This paper presents the results of many estimates of the performance impact of playing with Jose Canseco. None of the regression specifications employed identify a statistically significant and positive performance effect from playing with Canseco. There are numerous reasons why results reported in this paper do not support the findings of Gould and Kaplan; however, great care has been taken to ensure that the best possible specification choices were made. In light of these findings, it seems that the reported estimated Canseco spillover effects were likely the result of multiple flaws in the study's design. The performance record of Canseco's teammates does not provide evidence of a performance boost that followed Canseco as he moved from team to team.

The fact that a Canseco effect is not evident does not mean that baseball players did not use steroids or that Canseco did not aid his teammates with in using steroids. There is strong evidence that many baseball players, including former associates of Canseco, used performance-enhancing drugs. What the results do indicate is that being Jose Canseco's teammate did not provide any marginal performance advantage over players who were never teammates with Canseco. Thus, Canseco's role in spreading steroid use in baseball is likely exaggerated. In fact, the lack of results should not be surprising even in a world of rampant steroid use. The knowledge of how to use steroids is understood widely throughout sports among athletes, trainers, coaches, and doctors. Such dispersion of knowledge means that Canseco's influence would not be unique; thus, even if one accepts that Jose Canseco is being truthful about his involvement in disseminating steroid information, in a setting of pervasive steroid knowledge this behavior would not be identifiable using the chosen empirical strategy.

TABLE 4. The estimated peer effects of Jose Canseco on home runs per at-bat

	1	2	3	4	5	6	7	8	9	10
With Canseco	−0.00185* (0.00081)	−0.00122 (0.00103)	−0.00130 (0.00081)	−0.00052 (0.00107)	−0.00176* (0.00086)	−0.00253* (0.00113)	−0.00136 (0.00085)	−0.00255* (0.00120)	−0.00176* (0.00088)	−0.00244* (0.00120)
After Canseco	−0.00066 (0.00056)	0.00033 (0.00108)	0.00015 (0.00055)	−0.00831** (0.00145)	−0.0007 (0.00055)	0.00005 (0.00119)	0.00027 (0.00055)	−0.01358** (0.00286)	−0.00047 (0.00063)	−0.00046 (0.00132)
Career HR/AB	0.93474** (0.00763)		0.94597** (0.00759)		0.97408** (0.00791)		0.98818** (0.00779)		0.98519** (0.01076)	
Ballpark Factor	0.00015** (0.00002)	0.00010** (0.00003)	0.00015** (0.00002)	0.00075** (0.00008)	0.00017** (0.00002)	0.00013** (0.00004)	0.00017** (0.00002)	0.00088** (0.00014)	0.00016** (0.00003)	0.00011* (0.00005)
League R/G	0.00441** (0.00029)	0.00719** (0.00045)	0.00113 (0.00061)	0.00374** (0.00097)	0.00464** (0.00030)	0.00836** (0.00050)	0.00140* (0.00064)	−0.02106** (0.00434)	0.00497** (0.00042)	0.00770** (0.00068)
Age	0.00503** (0.00030)	−0.00016 (0.00025)	0.00519** (0.00030)	0.01422** (0.00144)	0.00439** (0.00033)	−0.00059* (0.00028)	0.00466** (0.00032)	0.01721** (0.00304)	0.00495** (0.00046)	−0.00062 (0.00039)
Age^2	−0.0000829** (0.0000051)	−0.0000041 (0.0000042)	−0.0000854** (0.0000050)	−0.0003917** (0.0000407)	−0.0000715** (0.0000055)	0.0000040 (0.0000047)	−0.0000757** (0.0000054)	−0.0004112** (0.0000732)	−0.0000776** (0.0000075)	0.0000062 (0.0000065)
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year Effects	No	No	Yes	Yes	No	No	Yes	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	50	200	200	200	200	200	200
Observations	16,746	13,752	16,746	13,752	11,551	9,402	11,551	9,402	6,612	5,266
Players	2,994	2,418	2,994	2,418	2,149	1,724	2,149	1,724	1,346	1,081
R^2 (overall)	0.63	0.03	0.64	0.00	0.72	0.04	0.73	0.00	0.71	0.04
R^2 (within)	0.07	0.05	0.09	0.00	0.08	0.07	0.11	0.00	0.09	0.06

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 5. The estimated peer effects of Jose Canseco on home runs per hit

	1	2	3	5	6	7	9	10
With Canseco	−0.00619* (0.00310)	−0.00291 (0.00390)	−0.00469 (0.00310)	−0.00590 (0.00306)	−0.00758 (0.00402)	−0.00477 (0.00304)	−0.00607 (0.00313)	−0.00728 (0.00421)
After Canseco	−0.00141 (0.00213)	0.00178 (0.00409)	0.00125 (0.00213)	−0.00183 (0.00197)	0.00004 (0.00421)	0.00127 (0.00195)	−0.00083 (0.00221)	−0.00036 (0.00458)
Career HR/H	0.94916** (0.00765)		0.95693** (0.00763)	0.98108** (0.00753)		0.99105** (0.00743)	0.98820** (0.01013)	
Ballpark Factor	0.00043** (0.00009)	0.00031** (0.00012)	0.00042** (0.00009)	0.00049** (0.00009)	0.00039** (0.00013)	0.00049** (0.00009)	0.00039** (0.00012)	0.00024 (0.00017)
League R/G	0.01385** (0.00110)	0.02295** (0.00170)	0.00432 (0.00235)	0.01379** (0.00109)	0.02511** (0.00178)	0.00487* (0.00230)	0.01419** (0.00149)	0.02194** (0.00240)
Age	0.01434** (0.00116)	−0.00011 (0.00096)	0.01492** (0.00115)	0.01227** (0.00118)	−0.00126 (0.00100)	0.01319** (0.00116)	0.01418** (0.00161)	−0.00044 (0.00137)
Age^2	−0.0002314** (0.0000193)	−0.0000088 (0.0000159)	−0.0002405** (0.0000191)	−0.0001948** (0.0000196)	0.0000137 (0.0000166)	−0.0002090** (0.0000192)	−0.0002165** (0.0000264)	0.0000052 (0.0000229)
Fixed Effects	No	Yes	No	No	Yes	No	No	Yes
Year Effects	No	No	Yes	No	No	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	200	200	200	200	200
Observations	16,746	13,752	16,746	11,551	9,402	11,551	6,612	5,266
Players	2,994	2,418	2,994	2,149	1,724	2,149	1,346	1,081
R^2 (overall)	0.63	0.03	0.64	0.73	0.04	0.74	0.73	0.03
R^2 (within)	0.05	0.04	0.06	0.06	0.06	0.09	0.07	0.05

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 6. The estimated peer effects of Jose Canseco on extra-base hits per at-bat

	1	2	3	4	5	6	7	9	10
With Canseco	−0.00391** (0.00136)	−0.00416* (0.00171)	−0.00263 (0.00136)	0.02049 (0.05616)	−0.00412** (0.00135)	−0.00486** (0.00176)	−0.00318* (0.00134)	−0.00408** (0.00139)	
After Canseco	−0.00153 (0.00092)	−0.00011 (0.00177)	−0.00041 (0.00092)	0.01091 (0.02476)	−0.00095 (0.00085)	−0.00081 (0.00182)	0.00027 (0.00084)	−0.00062 (0.00096)	
Career XB/AB	0.92749** (0.00919)		0.94908** (0.00943)		0.94611** (0.00942)		0.97326** (0.00963)	0.96813** (0.01311)	
Ballpark Factor	0.00025** (0.00004)	0.00025** (0.00005)	0.00025** (0.00004)	−0.00024 (0.00121)	0.00028** (0.00004)	0.00032** (0.00006)	0.00027** (0.00004)	0.00029** (0.00005)	
League R/G	0.00759** (0.00049)	0.01183** (0.00074)	0.00152 (0.00102)	−0.00811 (0.03389)	0.00769** (0.00048)	0.01274** (0.00078)	0.00121 (0.00101)	0.00801** (0.00066)	
Age	0.00833** (0.00050)	0.00197** (0.00042)	0.00856** (0.00050)	−0.00435 (0.01720)	0.00694** (0.00051)	0.00154** (0.00044)	0.00729** (0.00050)	0.00778** (0.00070)	
Age^2	−0.0001411** (0.0000084)	−0.0000480** (0.0000070)	−0.0001447** (0.0000083)	0.0002221 (0.0006623)	−0.0001173** (0.0000085)	−0.0000390** (0.0000073)	−0.0001226** (0.0000083)	−0.0001260** (0.0000115)	
Fixed Effects	No	Yes	No	Yes	No	Yes	No	No	Yes
Year Effects	No	No	Yes	Yes	No	No	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	50	200	200	200	200	200
Observations	16,746	13,752	16,746	13,752	11,551	9,402	11,551	6,612	5,266
Players	2,994	2,418	2,994	2,418	2,149	1,724	2,149	1,346	1,081
R^2 (overall)	0.54	0.05	0.55	0.01	0.63	0.07	0.64	0.61	0.07
R^2 (within)	0.07	0.09	0.08	0.00	0.08	0.13	0.1	0.08	0.12

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 7. The estimated peer effects of Jose Canseco on extra-base hits per hit

	1	2	3	5	6	7	9	10
With Canseco	−0.01486** (0.00479)	−0.01559** (0.00598)	−0.01192* (0.00481)	−0.01429** (0.00445)	−0.01588** (0.00576)	−0.01207** (0.00445)	−0.01459** (0.00455)	−0.01531* (0.00603)
After Canseco	−0.00364 (0.00326)	−0.00186 (0.00622)	−0.00079 (0.00326)	−0.00181 (0.00281)	−0.00481 (0.00594)	0.00117 (0.00280)	−0.00065 (0.00312)	−0.00609 (0.00642)
Career XB/H	0.95033** (0.00919)		0.96359** (0.00948)	0.97115** (0.00881)		0.98736** (0.00902)	0.98605** (0.01207)	
Ballpark Factor	0.00047** (0.00014)	0.00061** (0.00018)	0.00048** (0.00014)	0.00059** (0.00013)	0.00087** (0.00018)	0.00059** (0.00012)	0.00055** (0.00017)	0.00070** (0.00024)
League R/G	0.01983** (0.00172)	0.03143** (0.00260)	0.00543 (0.00362)	0.01873** (0.00159)	0.03157** (0.00255)	0.00299 (0.00333)	0.01851** (0.00215)	0.02803** (0.00343)
Age	0.01871** (0.00177)	0.00937** (0.00146)	0.01929** (0.00176)	0.01532** (0.00168)	0.00777** (0.00143)	0.01618** (0.00166)	0.01726** (0.00227)	0.00784** (0.00196)
Age^2	−0.0003069** (0.0000296)	−0.0001690** (0.0000243)	−0.0003162** (0.0000294)	−0.0002513** (0.0000279)	−0.0001415** (0.0000238)	−0.0002644** (0.0000276)	−0.0002699** (0.0000373)	−0.0001297** (0.0000327)
Fixed Effects	No	Yes	No	No	Yes	No	No	Yes
Year Effects	No	No	Yes	No	No	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	200	200	200	200	200
Observations	16,746	13,752	16,746	11,551	9,402	11,551	6,612	5,266
Players	2,994	2,418	2,994	2,149	1,724	2,149	1,346	1,081
R^2 (overall)	0.54	0.06	0.55	0.66	0.07	0.67	0.64	0.05
R^2 (within)	0.03	0.08	0.05	0.04	0.12	0.06	0.05	0.11

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 8. The estimated peer effects of Jose Canseco on isolated power

	1	2	3	4	5	6	7	8	9	10
With Canseco	−0.00762** (0.00276)	−0.00677 (0.00350)	−0.00509 (0.00275)	−0.02704** (0.00569)	−0.00740* (0.00287)	−0.00960* (0.00377)	−0.00559* (0.00284)	−0.00662 (0.00374)	−0.00725* (0.00295)	−0.00910* (0.00398)
After Canseco	−0.00295 (0.00189)	0.00049 (0.00366)	−0.00012 (0.00188)	−0.09539** (0.01873)	−0.00253 (0.00185)	−0.00110 (0.00394)	0.00069 (0.00182)	−0.00434 (0.00393)	−0.00179 (0.00209)	−0.00319 (0.00435)
Career ISO	0.93067** (0.00806)		0.94765** (0.00808)		0.96093** (0.00839)		0.98187** (0.00833)		0.97938** (0.01154)	
Ballpark Factor	0.00058** (0.00008)	0.00052** (0.00011)	0.00058** (0.00008)	0.00702** (0.00123)	0.00067** (0.00008)	0.00068** (0.00012)	0.00066** (0.00008)	0.00093** (0.00012)	0.00066** (0.00011)	0.00061** (0.00016)
League R/G	0.01684** (0.00099)	0.02676** (0.00152)	0.00345 (0.00208)	−0.29635** (0.05792)	0.01741** (0.00103)	0.02996** (0.00167)	0.00372 (0.00215)	0.00365 (0.00318)	0.01822** (0.00141)	0.02789** (0.00227)
Age	0.01832** (0.00103)	0.00183* (0.00086)	0.01888** (0.00102)	0.14680** (0.02679)	0.01549** (0.00111)	0.00044 (0.00094)	0.01638** (0.00108)	0.00980** (0.00122)	0.01741** (0.00152)	0.00012 (0.00130)
Age^2	−0.0003100** (0.0000172)	−0.0000653** (0.0000143)	−0.0003187** (0.0000170)	−0.0032993** (0.0006036)	−0.0002607** (0.0000184)	−0.0000382* (0.0000156)	−0.0002743** (0.0000179)	−0.0002163** (0.0000240)	−0.0002808** (0.0000251)	−0.0000252 (0.0000217)
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year Effects	No	No	Yes	Yes	No	No	Yes	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	50	200	200	200	200	200	200
Observations	16,746	13,752	16,746	13,752	11,551	9,402	11,551	9,402	6,612	5,266
Players	2,994	2,418	2,994	2,418	2,149	1,724	2,149	1,724	1,346	1,081
R^2 (overall)	0.61	0.04	0.62	0.00	0.69	0.05	0.70	0.00	0.68	0.05
R^2 (within)	0.08	0.08	0.10	0.00	0.09	0.12	0.12	0.15	0.10	0.11

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 9. The estimated peer effects of Jose Canseco on slugging average

	1	2	3	4	5	6	7	8	9	10
With Canseco	−0.00892* (0.00424)	−0.00699 (0.00531)	−0.00473 (0.00424)	−0.00541 (0.09654)	−0.00906* (0.00412)	−0.01067* (0.00535)	−0.00601 (0.00408)	−0.00607 (0.00534)	−0.00871* (0.00423)	−0.01009 (0.00568)
After Canseco	−0.00473 (0.00288)	0.00310 (0.00551)	−0.00049 (0.00287)	0.00357 (0.06515)	−0.00404 (0.00260)	0.00061 (0.00551)	0.00053 (0.00256)	−0.00315 (0.00553)	−0.00311 (0.00293)	−0.00341 (0.00609)
Career SLG	0.92921** (0.00964)		0.95052** (0.00970)		0.93430** (0.00984)		0.96214** (0.00983)		0.95388** (0.01363)	
Ballpark Factor	0.00096** (0.00012)	0.00114** (0.00016)	0.00095** (0.00012)	0.00117 (0.00191)	0.00102** (0.00012)	0.00129** (0.00017)	0.00100** (0.00011)	0.00150** (0.00017)	0.00106** (0.00016)	0.00119** (0.00023)
League R/G	0.02460** (0.00151)	0.04110** (0.00231)	0.00520 (0.00319)	0.03381 (0.04706)	0.02491** (0.00146)	0.04506** (0.00237)	0.00511 (0.00305)	0.00945* (0.00455)	0.02685** (0.00200)	0.04298** (0.00323)
Age	0.02900** (0.00157)	0.01337** (0.00130)	0.02984** (0.00155)	0.01454 (0.03754)	0.02367** (0.00156)	0.01230** (0.00133)	0.02492** (0.00153)	0.02150** (0.00172)	0.02643** (0.00214)	0.01186** (0.00185)
Age^2	−0.0004994** (0.0000262)	−0.0002849** (0.0000216)	−0.0005126** (0.0000260)	−0.0002976 (0.0009918)	−0.0004050** (0.0000259)	−0.0002531** (0.0000221)	−0.0004243** (0.0000254)	−0.0004040** (0.0000341)	−0.0004348** (0.0000352)	−0.0002350** (0.0000308)
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year Effects	No	No	Yes	Yes	No	No	Yes	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	50	200	200	200	200	200	200
Observations	16,746	13,752	16,746	13,752	11,551	9,402	11,551	9,402	6,612	5,266
Players	2,994	2,418	2,994	2,418	2,149	1,724	2,149	1,724	1,346	1,081
R^2 (overall)	0.52	0.03	0.53	0.00	0.60	0.05	0.62	0.01	0.59	0.05
R^2 (within)	0.08	0.17	0.10	0.17	0.09	0.24	0.12	0.26	0.10	0.24

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 10. The estimated peer effects of Jose Canseco on batting average

	1	2	3	4	5	6	7	8	9	10
With Canseco	−0.00126 (0.00221)	−0.00024 (0.00276)	0.00037 (0.00223)	0.00188 (0.00288)	−0.00144 (0.00197)	−0.00084 (0.00255)	−0.00028 (0.00198)	−0.00030 (0.00268)	−0.00124 (0.00200)	−0.00060 (0.00267)
After Canseco	−0.00198 (0.00148)	0.00251 (0.00282)	−0.00048 (0.00148)	0.00806 (0.00587)	−0.00183 (0.00120)	0.00151 (0.00256)	−0.00031 (0.00120)	0.00377 (0.00291)	−0.00161 (0.00133)	−0.00023 (0.00277)
Career AVG	0.96778** (0.01226)		0.97744** (0.01226)		0.93566** (0.01258)		0.95186** (0.01255)		0.92846** (0.01711)	
Ballpark Factor	0.00037** (0.00006)	0.00063** (0.00008)	0.00037** (0.00006)	0.00015 (0.00044)	0.00035** (0.00005)	0.00065** (0.00008)	0.00034** (0.00005)	0.00032* (0.00015)	0.0004** (0.00007)	0.00062** (0.00011)
League R/G	0.00754** (0.00077)	0.01454** (0.00120)	0.00173 (0.00167)	−0.00473 (0.00727)	0.00697** (0.00068)	0.01549** (0.00114)	0.00116 (0.00146)	0.02161** (0.00502)	0.00835** (0.00092)	0.01545** (0.00151)
Age	0.01089** (0.00081)	0.01133** (0.00067)	0.01108** (0.00080)	0.00341 (0.00801)	0.00843** (0.00072)	0.01141** (0.00064)	0.00869** (0.00072)	0.00566* (0.00276)	0.00911** (0.00097)	0.01135** (0.00088)
Age^2	−0.0001936** (0.0000135)	−0.0002162** (0.0000112)	−0.0001962** (0.0000134)	0.0000592 (0.0002424)	−0.0001488** (0.0000120)	−0.0002082** (0.0000106)	−0.0001525** (0.0000119)	−0.0000545 (0.0000656)	−0.0001551** (0.0000160)	−0.0002043** (0.0000146)
Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Year Effects	No	No	Yes	Yes	No	No	Yes	Yes	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1982-2004	1982-2004
PA Cutoff	50	50	50	50	200	200	200	200	200	200
Observations	16,746	13,752	16,746	13,752	11,551	9,402	11,551	9,402	6,612	5,266
Players	2,994	2,418	2,994	2,418	2,149	1,724	2,149	1,724	1,346	1,081
R^2 (overall)	0.38	0.01	0.39	0.00	0.43	0.02	0.45	0.00	0.42	0.02
R^2 (within)	0.05	0.19	0.05	0.13	0.05	0.27	0.06	0.23	0.05	0.27

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

TABLE 11. The estimated peer effects of Jose Canseco on specifically-identified teammates

	HR/AB	HR/H	XB/AB	XB/H	ISO	SLG	AVG
With Canseco	−0.00313 (0.00514)	−0.01223 (0.01672)	0.00164 (0.00658)	0.00676 (0.01934)	−0.00175 (0.01601)	−0.00296 (0.02122)	−0.00001 (0.00817)
After Canseco	−0.00276 (0.00587)	−0.01247 (0.02078)	−0.00136 (0.00672)	−0.00971 (0.02094)	−0.00882 (0.01800)	−0.00467 (0.02209)	−0.00257 (0.00812)
Career Statistic for Dependent Variable	1.07506** (0.11632)	1.10397** (0.11662)	0.95045** (0.14088)	1.01605** (0.11779)	1.0483** (0.12899)	0.97931** (0.15858)	0.76613** (0.25823)
Ballpark Factor	0.00041 (0.00043)	0.00017 (0.00142)	0.00097 (0.00054)	0.00147 (0.00160)	0.00174 (0.00134)	0.00301 (0.00175)	0.00133 (0.00071)
League R/G	0.01451** (0.00540)	0.03142 (0.01793)	0.01684* (0.00668)	0.03452 (0.01995)	0.04518** (0.01676)	0.06415** (0.02169)	0.01728* (0.00864)
Age	0.00988 (0.00519)	0.02095 (0.01885)	0.01836** (0.00580)	0.03203 (0.01820)	0.03981* (0.01586)	0.06222** (0.01918)	0.02587** (0.00757)
Age^2	−0.0001551 (0.0000846)	−0.0003013 (0.0003074)	−0.0003049** (0.0000946)	−0.0005161 (0.0002968)	−0.0006443* (0.0002587)	−0.0010241** (0.0003127)	−0.000436** (0.0001235)
Fixed Effects	No	No	No	No	No	No	No
Year Effects	No	No	No	No	No	No	No
Sample Period	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010	1970-2010
PA Cutoff	200	200	200	200	200	200	200
Observations	113	113	113	113	113	113	113
Players	8	8	8	8	8	8	8
R^2 (overall)	0.68	0.72	0.54	0.62	0.63	0.53	0.39
R^2 (within)	0.28	0.20	0.28	0.14	0.28	0.33	0.26

Standard errors in parentheses; * significant at 5%; ** significant at 1%.

References

Baltagi, Badi H., and Ping X. Wu. 1999. Unequally Spaced Panel Data Regressions with AR(1) Disturbances. *Econometric Theory* 15(6): 814-823.

Bradbury, John Charles. 2008. Was Jose Canseco the Johnny Appleseed of Steroids? *Sabernomics: Economic Thinking About Baseball*, February 15. **Link**

Bradbury, John Charles. 2009. Peak Athletic Performance and Ageing: Evidence from Baseball. *Journal of Sports Sciences* 27(6): 599-610.

Bradbury, John Charles. 2010a. *Hot Stove Economics: Understanding Baseball's Second Season*. New York: Copernicus.

Bradbury, John Charles. 2010b. Hired to Be Fired: The Publicity Value of Managers. Working paper. **Link**

Bradbury, John Charles, and Douglas J. Drinen. 2008. Pigou at the Plate: Externalities in Major League Baseball. *Journal of Sports Economics* 9(2): 211-224.

Canseco, Jose. 2005. *Juiced: Wild Times, Rampant 'Roids, Smash Hits, and How Baseball Got Big*. New York: Regan Books.

Canseco, Jose. 2008. *Vindicated: Big Names, Big Liars, and the Battle to Save Baseball*. New York: Simon Spotlight Entertainment.

De Vany, Arthur. 2011. Steroids and Home Runs. *Economic Inquiry* 49(2): 489-511.

Fisman, Ray. 2010. The Canseco Effect: Wherever He Went, Jose Canseco Made His Teammates Better Power Hitters. Can Statistics Be Used to Find Juicers? *Slate*, July 30. **Link**

Freakonomics. 2010a. The Canseco Effect? *Freakonomics Blog*, August 3. **Link**

Freakonomics. 2010b. Disputing the Canseco Effect. *Freakonomics Blog*, August 5. **Link**

Gould, Eric D., and Todd R. Kaplan. 2011. Learning Unethical Practices from a Co-Worker: The Peer Effect of Jose Canseco. *Labour Economics* 18(3): 338-348.

Hartgens, Fred, and Harm Kuipers. 2004. Effects of Androgenic-Anabolic Steroids in Athletes. *Sports Medicine* 34(8): 513-554.

James, Bill. *The 1982 Bill James Abstract*. 1982. New York: Ballantine Books.

Merkin, Scott. 2003. Amid Controversy, Sox Take Tests: Union Rep Wunsch Confirms All Players Tested. *MLB.com*, March 11. **Link**

Schell, Michael J. 2005. *Baseball's All-Time Best Sluggers: Adjusted Batting Performance from Strikeouts to Home Runs*. Princeton, N.J.: Princeton University Press.

Schulz, Richard, and Christine Curnow. 1988. Peak Performance and Age Among Superathletes: Track and Field, Swimming Baseball, Tennis, and Golf. *Journal of Gerontology* 43(5): 113-120.

Stacy, Jason J., Thomas R. Terrell, and Thomas D. Armsey. 2004. Ergogenic Aids: Human Growth Hormone. *Current Sports Medicine Reports* 3(4): 229-233.

Tobin, Roger G. 2008. On the Potential of a Chemical Bonds: Possible Effects of Steroids on Home Run Production in Baseball. *American Journal of Physics* 76(1): 15-20.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

About the Author

J. C. Bradbury is professor of sport management and economics at Kennesaw State University. He is the author of *Hot Stove Economics: Understanding Baseball's Second Season* and *The Baseball Economist: The Real Game Exposed*. His email address is jcbradbury@kennesaw.edu.

Eric Gould and Todd Kaplan's reply to this article
Go to Archive of Comments section
Go to January 2013 issue

Discuss this article at Journaltalk:
http://journaltalk.net/articles/5788