

Sports Analytics: A View from Behind the Foul (Fair) Pole

Rick L. Wilson
Oklahoma State University
7/26/2011
rick.wilson@okstate.edu

Talk Outline

- Background Information/Motivation
- Sports Analytics Impact Levels
- Selected Sports Analytics observations
- Future Analytics?
- Final anecdotes about BCS, MLB Realignment, and lessons learned

Disclaimer

- Any opinions expressed here are mine and no one else is responsible. If necessary though, I'll blame BYU and Howard Schnellenburger (circa 1984), Dr. John Gleason and Dr. Dick Overfield (circa 1985), the owners and players that cancelled the 1994 World Series, among others, for any misguided ideas or thoughts.

Why Sports Analytics?

- Link to Movie trailer

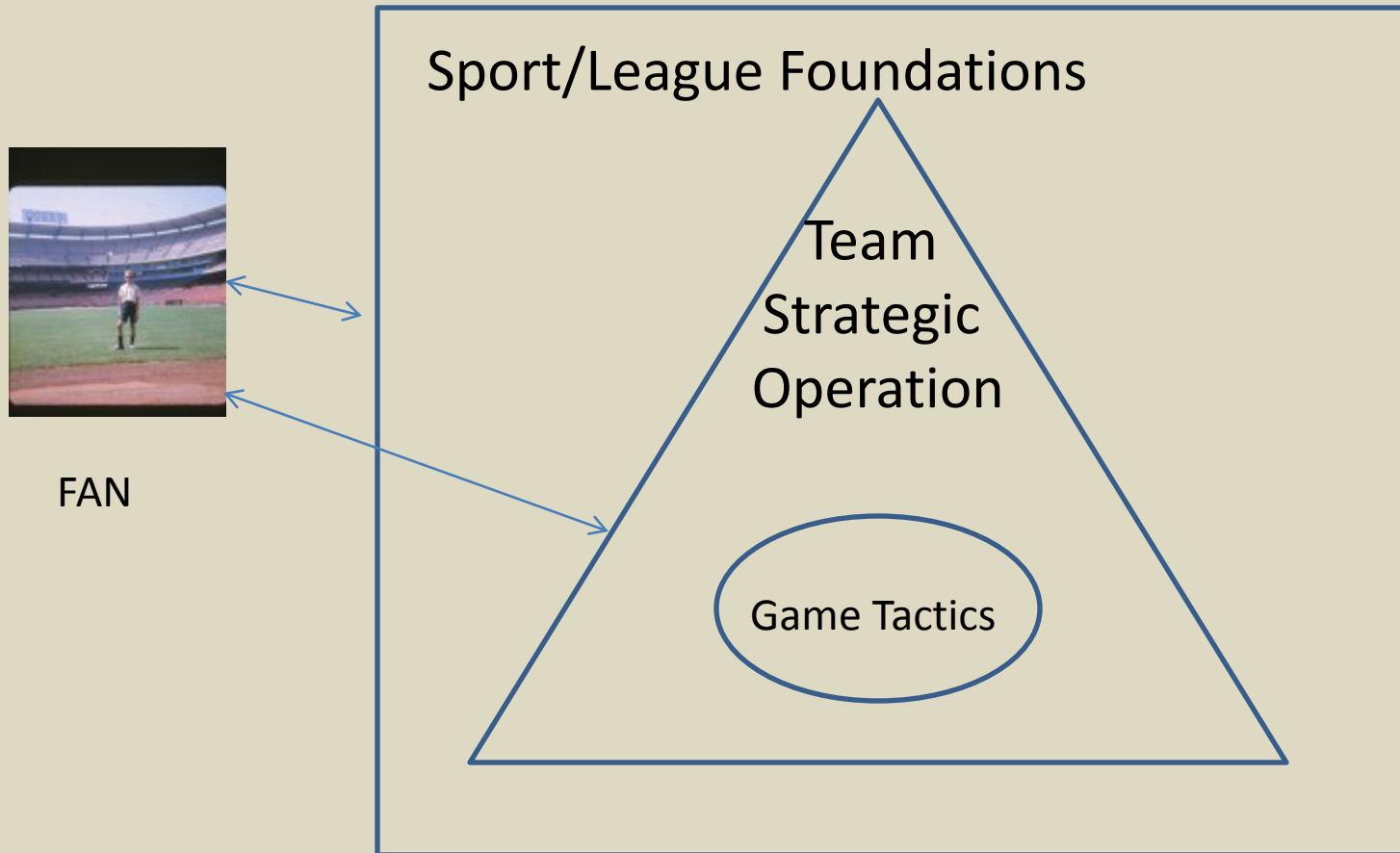
Why Sports Analytics?

- 2010-2011 NBA Champions – Dallas Mavericks
 - Owner – Mark Cuban
- Our on-going love affair with ‘fantasy’ sports and our fascination with data
- No one would be interested in my research study on the moderating impact of math anxiety on on-line operations management students.

Framework for Sports Analytics

- Analytics Definition
 - Statistics
 - Operations research/management science
 - Data Mining
 - Knowledge Management (organizational aspects of quantitative and qualitative analysis).
- Decision Types
 - Sports Foundations
 - Strategic (Organization)
 - Tactical (Play of sport)

Sports Analytic Lens



Sports Analytics

- Goals
 - Foundational – better management or direction of the ‘league’ or sport
 - Strategic – better ‘team’
 - Tactical – better execution during the game
- Teams/Fans/Observers
 - Identify new found insights about some phenomenon.
 - Improve Fan experience through data access

Baseball - Example of Different Levels

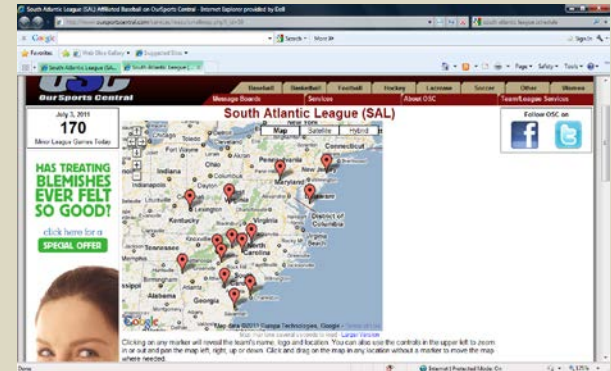
- Foundation – how baseball is scheduled
- Strategic
 - Valuation/selection of players
 - Forecasting player performance
 - Drafting college players vs. HS
 - Stage in career/free agency
- Tactical
 - Batting order (LaRussa pitcher 8th)
 - Stealing vs. Waiting for the big inning

Baseball Analytics – Foundation Level

- Baseball Scheduling
 - Incredibly difficult – huge mathematical problem
 - ? What about realignment?
 - Husband and Wife team for many years
 - OR guru Mike Trick – couple of years used sophisticated IP techniques
 - Interleague play
 - “Repeaters”
- Umpire Scheduling
 - Example from my class this summer

Sally League Scheduling – Exam Question

- Games played on July 6th – (Umpire crew designation also shown).
- CREW 1 - [Charleston RiverDogs](#) at [Rome Braves](#)
- CREW 2 - [Greensboro Grasshoppers](#) at [West Virginia Power](#)
- CREW 3 - [Lakewood BlueClaws](#) at [Delmarva Shorebirds](#)
- CREW 4 - [Hickory Crawdads](#) at [Greenville Drive](#)
- CREW 5 - [Savannah Sand Gnats](#) at [Augusta GreenJackets](#)
- CREW 6 - [Lexington Legends](#) at [Hagerstown Suns](#)
- CREW 7 - [Asheville Tourists](#) at [Kannapolis Intimidators](#)



- Series(games) starting on July 8th . Umpires cannot repeat teams.
- July 8, 2012
- [Lexington Legends](#) at [Hickory Crawdads](#)
- [Greenville Drive](#) at [Rome Braves](#)
- [Augusta GreenJackets](#) at [Asheville Tourists](#)
- [Delmarva Shorebirds](#) at [Charleston RiverDogs](#)
- [West Virginia Power](#) at [Hagerstown Suns](#)
- [Kannapolis Intimidators](#) at [Lakewood BlueClaws](#)
- [Greensboro Grasshoppers](#) at [Savannah Sand Gnats](#)

	Rome GA	Hickory NC	Asheville NC	Charleston SC	Hagerstown MD	Lakewood NJ	Savannah GA
Rome GA	0	296	221	361	636	851	318
Charleston WV	476	251	292	471	293	514	515
Salisbury MD	778	463	537	570	124	192	613
Greenville SC	202	115	63	213	499	690	257
Augusta GA	216	215	182	143	568	745	135
Hagerstown MD	636	394	452	599	0	234	643
Kannapolis NC	322	63	137	234	380	570	277

Sampling of Sports - Data

- Baseball – statistics galore! Individual based team
- Basketball – better than football, sufficient data, unique team dependency
- Football – less data, need unique game film, more team dependency.
- Hockey – infancy, unique team dependency
- Soccer – similar to Hockey, even less ‘fan friendly’ data.
- Golf – Individual sport – no general managers

Lead Sport - Baseball – Key Dates

- Bill James – SABR – Society for American Baseball Research (now consultant for Boston Red Sox) – writing in 1970’s-80’s.
- Billy Beane – “MoneyBall” – Oakland “A’s” – Michael Lewis book 2003 – Movie 9/2011.
- Rotisserie Baseball – 1984
- October 29, 1969 – UCLA-Stanford Lab experiment (“LOG..”) – Data availability

Baseball's Analytic Measures

Past Data

- Pitching
 - Win/Losses
 - ERA
 - Saves (after 196x)
- Batting
 - Avg.
 - HR
 - RBI
 - SB

Present/Future Data

- Win Shares
- Peripheral ERA (PERA)
- DIPS (Def. Ind. Pitching Stat)
 - $(13*HR + 3*BB - 2*K)/IP + 3.7$
- Runs Created
- OPS = OBP + SLG
- VORP (Value over repl. Player)

Overview of Findings

- Runs created key
- DIPS – a measure of pitchers effectiveness
- Projections – PECOTA – crossover from fantasy?
- Scouting plus numbers – strategic use
- Tactics about stealing, bunting, hit and run – different?
- Drafting different (HS vs. College, ROE).

Baseball Analytics Today

- Everyone has an operations 'staff'
- Tension between fantasy analysts and 'baseball' analysts
- Some attribute today's parity in baseball to the fact that 'everyone has access to data and analysis tools'.
 - Close pennant races
 - Lack of 100 game winners

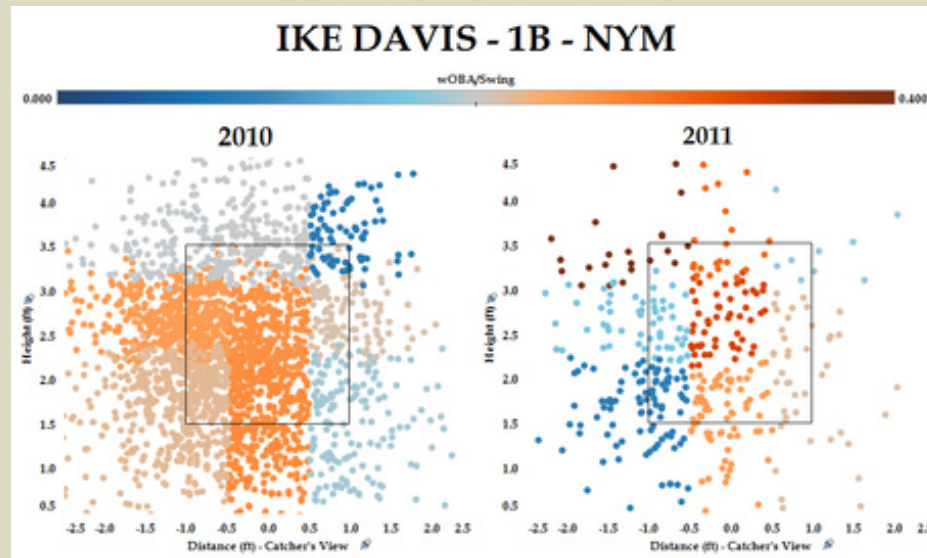
Baseball Analysis Today

- My experience with Colorado Rockies
 - Incredible Data, Video resources, commitment
 - First to do video on iPods for players, etc.
 - Brian Jones, Jimmy Hartley – OSU grads (not in picture).



Trends in Baseball Analytics

- More complicated measures
 - Makes my head spin!
- f/x for pitching/hitting – even more data?



Data in the 'Wrong' Hands – the Press

- Attendance – “Baseball Attendance Down – 18 of 30 teams shows decrease!” – Forbes, July 2011. (Data shows 17 of 30, and a net increase of +59/team for 2011).
- “A team’s run differential is normally a better indication of its level of play than its record at the halfway point...” ??? – SI, 7/18/2011.

Data in the 'Wrong' Hands – the Press

- July 10th – Yahoo Sports – AccuScore – “Home Run Derby curse is real” – 24 of 40 finalists/semi-finalists saw a decrease in SLGPCT in 2nd half of season.
 - First, it’s not statistically significant (meaning – it can easily happen by chance).
 - Second, participants probably had a great 1st half – that’s why they were invited. Makes sense for 2nd half decrease.
- **BE CAREFUL WITH YOUR DATA!!**

Basketball

- Mark Cuban – first owner to embrace Analytics
 - MIT conference attendee (more later)
- Dallas Mavericks 2010-2011 Champions!
- Many teams employ ops people in NBA.
- 82games.com founder works for Mavericks now – sits on the bench during games (Roland Beech).
- Metrics surround team performance beyond Pts, Reb., Assists. (Plus/Minus).

Hockey/Soccer

- Not much to quantify – Red Wings GM
- Phil Scarf in UK – involved with research in Soccer Player Evaluations
- Previous research in penalty kick order (relevant after Woman's World Cup?)

Football – NFL and College

- NFL very secretive, need detailed films to isolate on players unlike baseball, basketball.
- Recently used analytics to modify OT system
 - Coin toss winners won disproportionate amount of the time.
- NFL, College and High School football (and other sports including basketball, wrestling, etc.) use HUDL System developed at University of Nebraska – www.hudl.com

Costs: \$800/yr for HSchool

hudl Contact Us: (402) 817-0060

FEATURES PARTNERS BUZZ BLOG ABOUT SUPPORT & HELP PRICING & SIGN-UP

See it in action
Click to watch a video

Hudl allows our basketball team to play smarter. We've used other systems out there and Hudl has continuously proven to be the most user friendly. It allows our coaching staff to teach the sport instead of just coach it.

Jon Youngblood
Head Basketball Coach
Southside Central High School

Every time I have called the Hudl support team they have answered on the first or second ring. The staff is extremely knowledgeable and makes every effort to answer all of my questions.

From DVD scene detection to unlimited coaching stations - Hudl is always ahead of the curve.

Nick Restifo
Assistant Football Coach
St. Ignace HS Wichita

Mike Johnson

hudl Varsity Football Dashboard Highlights Library Manage Coach Kalls Help

2016-2017 Season
2015-2016 Season
2016-2015 Season
2013-2014 Season
2012-2013 Season
2011-2012 Season
2010-2011 Season
2009-2010 Season
2008-2009 Season
Fall Camp Aug 18
Jacksonville Sep 3
KIA Nov 12
Central Sep 26
Southeast Oct 3
Northwest Oct 17
Watch Footage
Practice
Management Setup (3)
Northwest vs Central 11/28/2009

CLIP #	PLAY #	COV	SN	SECT	TEAM	YARD LN	PLAY TYPE	RESULT	DR/LB	OFF FORM	OFF PLAY	OFF STR	PLAY DIR	GAP	PASS CONC	SET PREVLT	COVERAGE
1	1	F	1	10	L	-28	Run	Rush	2	BUNCH TITE GUN	14 ZONE READ	R	L	A	4-3 SUP	3	S
2	2	D	2	8	L	-30	Run	Out of Bounds	45		MESH 14 HIGH	R	L	B	4-3 FUP	3	S
3	3	D	1	5	L	-25	Run	Back	25	ACE OVER	28 SWEEP	R	R	D	4-3 FUP	3	S
4	4	E	0	3	M	3	Extra Pt.	Good									
5	5	E	0	0	W	-40	Return	Return		DOUBLE WING							
6	6	E	0	1	R	-47	Run	Rush	1		15 TRIPLE OPTION	L	L	D	3-5	0	A
7	7	D	2	9	L	-48	Run	1st DN	29	DASH-TEST03-WORK	18 LEAD OPTION	R	R	D	3-5	3	A
8	8	D	1	4	L	-48	Run	Rush			18 LEAD OPTION	R	R	D	3-5	3	A

Tactical Strategies: 4th Down Conventional Wisdom

- Kevin Kelley – Pulaski Academy in Arkansas
 - Never punts, shows how $E(\text{points})$ is increased
 - Won state championship in 2009, 2nd in 2010.
- Bill Belichick – class example – Patriots vs. Colts, Nov 2009.
 - 4th and 2 – Decision Analysis would indicate he made the optimal decision to go for it.

College football – decision analysis

- 1984 Orange Bowl – 2 pt conversion
 - Decision Analysis shows that Dr. Osborne, by his decision, made poor decisions in 4th Quarter.
 - Think back to 1969
 - Texas vs. Arkansas

College football – OT Analysis

- Started in 1996 – possessions at the 25 yard line.
- TV Commentators, coaches “Huge Advantage”
- JQAS – Rosen and Wilson (2007)
 - found that most important factor was point spread
 - Some evidence that scoring TD first was best.
 - Conventional wisdom was not true (but wasn’t bad either).

Golf Analytics

- Individual Sport – no GM – any use for golfers?
- ShotLink – 7 million golf shots in database
- Traditional golf measures poor – Driving accuracy, Greens-In-Reg.

Golf analytics – “Shot Value”

- Dr. Broadie, Finance, Columbia University
- Fractional par value every place on hole based on scratch golfer
- Shot from point 1 (3.2), lands at point 2 (1.8).
- Shot value = $3.2 - 1.8 - 1 = .4$ BETTER than scratch.
- This analysis showed that shots from 200-250 are most important, 150-200 next. NOT PUTTING!

Sports Analytics and Future Research

- 1985 “I don’t like sports papers” – Dr. Gleason.
- 1990’s – you hid the fact you did sports research.
- 2011 – Out of the locker room (and no longer hiding behind the fair pole!).
- As an example, the 2011 MIT Sports Analytics Conference – summary of the topics. Headlined by Mark Cuban and others.

What did we learn?

- Teams feel they must keep up with the Beanes.
- Is this really helpful : (insert long equation here) – at what point does the research over complicate things?
- Data analyzed poorly can lead to new, wrong conventional wisdoms.

What did we learn?

- Is our organization is search of our own SABR-like metrics?
- Are we stuck on old conventional wisdom?
- Are we misinterpreting our data?
- Are we TOO focused on KPI and forgetting about the intangibles?

Future Analytic Help?

- Baseball Realignment
 - 30 teams, 15 each league.
 - Scheduling issue – players dislike inter-league games.
 - Simple plan would work:
 - No divisions
 - 12 interleague games (4x3), 150 within (8x12, 6x9)
 - Fit wonderfully in 27 weeks, 54 scheduling slots.
 - Could cluster teams, SOS comparable.

Future Analytic Help?

- Bowl Championship Series vs. Playoff
 - Trade-off
 - More teams – less likely to have ‘best team’ win.
 - How do you assess teams that should participate?
What is your objective?
 - Analytic approaches to compare teams based on SOUND theory and procedures date back to 1930’s. Why not use them? (See 1995 Interfaces article).

Conclusion

- We will be inundated with data even more in the future (sports, business).
- Goal of Analytics should be to help us bring clarity, understanding, and validate (or disprove) our conventional wisdoms.
- I'm glad its you in reality – but I'll try to do my part in the classroom for your future employees.