Why 0.400 Hitters Have Disappeared: 30-Year Evolution of the Korea Professional Baseball League

Jin Yeon Ahn[1] , Geun Ju Byun[2], Hyungho Byun[3], Young Chekal[4], Gwanyul Cheon[5], Byeong Hoon Cheon[6], Sung Haeng Cho[7], Shinhaeng Choi, Ku Hun Chung[7], Beom Hong[8], Soojin Jang[9], Won Cheol Jang[10], Sang Min Jeon[11], Jaeseung Jeong[12]*, Kihong Jun[13], Yongchin Jung[14], Minseung Kang[15], Ki Min Kim[16], Kisang Kim[17], Dae Joong Kim[18], Dong Sim Kim[19], Lee Yeon Kim[20], Sang Mo Kim[21], Seong Wan Kim[22], Yeon Jung Kim[23], Yong Nam Kim[24], Yoo Kyung Kim, Joo Hwan Kim[25], Tae Han Kim[26], Hyunjoo Kim[27], Hyo-Im Kim[28], Jong Heon Kwon[29], Kyujong Lee[30], Sun Hye Lee[31], Seul Ki Lee[32], Joung Seol Lee[33], Joon Soo Lee[34], Choong-Han Lee[35], Hyung Keuk Lee[36], Seonnam Lim[34], Sungsu Lim[37], Sanguk Nam[38], Seung Woo Nam[39], Nam Hee Noh[40], Jae mahn Noh[41], Wonky Oh[42], Mijoon Pak[43], Sang Hwa Park, Sungkul Park[44], Soohyun Park[45], Jonghyuck H. Park[46], Chan Eon Park[37], Hye Jung Park[47], Boo Gil Shin[48], EunKyo Shin[49], Eun Ju Song [50], Minho Yi[51], Shin-young Yoon[52]

[1]LUVE Inc., Seoul, South Korea
[2]Department of English Language & Literature, Kyonggi University, Suwon, South Korea
[3]Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
[4]Simplex Internet, Inc., Seoul, South Korea
[5]SISAIN magazine, Seoul, South Korea
[6]RealNetworks Asia Pacific Co., Ltd., Seoul, South Korea
[7]SAMOO Architects & Engineers, Seoul, South Korea
[8]Department of Physics, KAIST, Daejeon, South Korea
[9]Division of EcoScience, Ewha Womans University, Seoul, South Korea
[10]University of Georgia, United States of America
[11]Department of Chemical Engjneering, SungKyunKwan University, Seoul, South Korea
[12]Department of Bio and Brain Engineering, KAIST, Daejeon, South Korea
[13] SpaceBeing cafe, Seoul, South Korea
[14]Data Analysis Team, Baek In-Cheon Project
[15]Department of Energy Resources & Computer Science, Seoul National University, Seoul, South Korea
[16]Department of Economics, Korea University, Seoul, South Korea
[17]Korea Foundation for the Advancement of Science & Creativity, Seoul, South Korea
[18]KSA & IENTER, Seoul, South Korea
[19]Department of Educational technology, Ewha Womans University, Seoul, South Korea
[20]School of Medicine, Ewha Womans University, Seoul, South Korea
[21]IT Support Team, Baek In-Cheon Project
[22]School of Electrical & Electronic Engineering, Yonsei University, Seoul, South Korea
[23]Thinkware Systems Corporation, Seoul, South Korea
[24]Department of Educational Psychology, University of Wisconsin - Madison, United States of America
[25]Ministry of Public Administration and Security of the Republic of Korea, Seoul, South Korea
[26]TmaxSoft, Seoul, South Korea
[27]Department of Chemical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea
[28]School of Earth & Environmental Sciences, Seoul National University, Seoul, South Korea
[29]Department of Consumer & Family Science, SungKyunKwan University, Seoul, South Korea
[30]School of Industrial Management Engineering, Korea University, Seoul, South Korea
[31]Paxonline, Seoul, South Korea
[32]Department of Electrical Computer, University of Seoul, Seoul, South Korea
[33]Korea Institute of Science and Technology Information (KISTI), Daejeon, South Korea
[34]NC soft Co., Seoul, South Korea
[35]Department of Mechanical Engineering, Sungkyunkwan University
[36]Samsung Electronics Inc., Suwon, Korea
[37]Department of Mathematical Sciences, KAIST, Daejeon, South Korea

[38]Department of Economics, Seoul National University, Seoul, South Korea
[39]Hotel Prince Seoul, Seoul, South Korea
[40]Department of Archaeology and Art History, College of Humanities, Seoul National University, Seoul, South Korea
[41]Department of Tourism, Korea National Open Universtity, Seoul, South Korea
[42]Woongjin Ed Ltd., Seoul, South Korea
[43]Graduate School of Interpretation & Translation, Hankuk University of Foreign Studies, Seoul, South Korea
[44]Daewoo Securities Co., Ltd., Seoul, South Korea
[45]Jeonju District Court, Jeonju, South Korea
[46]Graduate School of Information Management and Security, Korea University
[47]Taylor Nelson Sofres Korea Ltd., Seoul, South Korea
[48]Department of Chemical Science Education, Graduate School of Education, Hanyang University, Seoul, South Korea
[49]Administration Team, Baek In-Cheon Project
[50]Department of Educational technology, Hanyang University, Seoul, South Korea
[51]Munhwa Broadcasting Corporation, Seoul, South Korea
[52]Donga Science magazine, Seoul, South Korea

*E-mail: jsjeong@kaist.ac.kr

ABSTRACT

The disappearance of batters with batting average of over 0.400 remained a mystery for a long time, until Stephen Jay Gould shed a light to this question by arguing that it is the result of overall improvement and stabilization of the Major League Baseball as a system rather than a decline in batter performance or enhanced pitcher abilities. Gould's major finding was that the standard deviation of the batting average tends to decline while the mean stays stable and this study is an attempt to apply Gould's theory to the Korean professional baseball league (KPB). A set of performance variables have been analyzed for a selective pool of batters, pitchers and to a certain extend fielders of the KPB for the last 30 years. The results show that the Gould's theory is effective in the KPB. The SD of batting averages of the KPB batters is showing statistically significant decline. The results of statistical analysis for other measurements are also largely supportive of this finding, especially in fielding performance measurements. Pitcher data showed mixed results, but at least we found that conventional explanations about deteriorating batter abilities or enhancement of pitching are not valid in the KPB according to the results of this study.

Keyword: Korea Professional Baseball league; disappearance of 0.400 hitters; Gould; In-Cheon Baek; collective intelligence

One of the most persistent baseball questions is about the disappearance of hitters with a batting average of over 0.400. The last seasonal batting average over 0.400 in the Major League Baseball (MLB) was achieved in 1941, well over a half century ago, by Ted Williams. Before him, the record was still a tremendous accomplishment, but not unheard of. There have been numerous attempts to explain the mysterious extinction of batting greatness. Conventional explanations range afar, from a nostalgic idealization of the past—Ted Williams himself, in his 1986 book *The Science of Hitting*, argued that there are simply not as many good hitters any more—to more scientific approaches to discovering an impediment to batting skills inherent to modern baseball—that the environment is tougher for batters nowadays.

In 1996, a new light has been shed to this question by Stephen Jay Gould[1]. Instead of trying to figure out the root cause behind the apparent deterioration of batting performance, he attacked the very foundation of the previous theories by arguing that the extinction of 0.400 batting evidenced improvement of the entire system of baseball, or what he dubbed the *full house*, rather than something of a slip. This rather counter-intuitive explanation is supported by the fact that since the 1920s, the standard deviation of the MLB batting average has steadily declined, while the seasonal batting averages remained stable. In other words, it was not just 0.400 hitters but also players at the bottom end of annual batting averages that went extinct, which is a sign of overall stabilization and improvement of competitive balance among batters. His theory has fascinated many and has been reinforced by many follow-up studies. There are studies which found that extreme performance is disappearing in baseball by analyzing the standard deviation (SD) of various performance measurements (Chatterjee and Hawkes, 1995; Leonard II, 1995; Miceli, 2005; and Schall and Smith, 2000). Indeed, Gould started a new tide of research in the study of baseball, which views the game from a holistic perspective instead of focusing on the behavior or ability of individual players. A study found how expansion of the player pool of the MLB sparked by immigration and global scouting efforts resulted in better competitive balance and homogeneity of skills among players (Schmidt and Berri, 2005). His approach was also applied to other sports, including basketball, football, soccer (Berry et al., 2005), tennis (Radicchi, 2011) and even to the stock market (Bernstein, 1998) and helped establish how overall quality of performance in different areas tended to enhance with time.

It presents an intriguing possibility to Korean baseball fans to consider what Gould's theory may hold for the Korean baseball scene. Throughout 30 years of history of the Korea Professional Baseball league (KPB)[2], the highest league of baseball in Korea, there was only one hitter who achieved the legendary 0.400 batting average. In-Cheon Baek of then MBC Blue Dragons hit

[1]Stephen Jay Gould, a Harvard-teaching American paleontologist, was also a renowned evolutionary biologist and known baseball fan.
[2]More information on the Korean professional baseball league is provided in the supplementary information (Table S1).

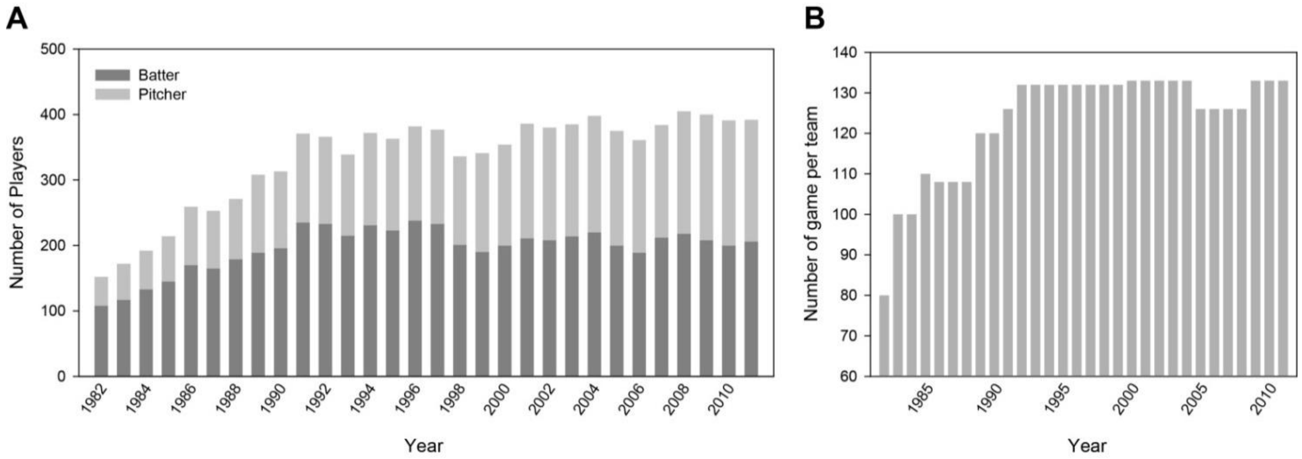

Figure 1 | (A) Total number of batters and pitchers in the data pool: Year 1982-2011
(B) Number of games per team: Year 1982-2011

0.412 in 1982, the very first year of the KPB. Since then, no one has ever reached the mark again. The closest records include Jong-beom Lee's 0.393 in 1994 and Hyo-jo Jang's 0.387 in 1987. Whenever a batter maintained mid-season batting average of over 0.400 for some time, fans and observers have been filled with hopeful speculations only to be disappointed later. Why has the record been so rare? There has been no serious attempt to answer this question in Korea. This study is an attempt to provide an answer by analyzing the 30-year data of the KPB game records so as to see whether Gould's theory will hold true in the Korean professional baseball.

It was not simple to conduct a study of Sabermetrics[3] in Korea. There are no organized research activities or full-time researcher of baseball statistics. To fulfill the prospect of applying Gould's theory to a league other than MLB for the first time, the initiator of this project made a proposal on a social network service to recruit individuals who might be interested in executing the project. The recruits represented diverse walks of life, including statistics students and professors, IT engineers, biology majors, journalists and of course, die-hard baseball fans. They were organized into six different functions: data collection, data analysis, thesis writing, visualization, IT support and administration. One of the biggest challenges inherent to a study of the KPB statistics was that the data from early years was still in the process of digitization and available only in a paper form, which meant that data collection and analysis would require a considerable amount of manual work. The data has been collected, digitized and cross-checked by all of the co-authors of this paper to ensure accuracy and credibility.

In this study, we are going to analyze various parameters of batting performance variables accumulated for the last 30 years with an aim to find out whether Gould's theory is still effective in the Korean professional baseball league. We also investigate pitching and to an extent possible fielding variables to explore the possibility of any additional support. By doing the above, we are going to consider the reasons why the KPB no longer sees hitters with a single-season batting average of 0.400.

Table 1 | Number and percentage of players in data set 2 (DS2)

Year	Number of pitchers			Number of batters		
	All	DS 2	DS 2 / All (%)	All	DS 2	DS 2 / All (%)
1982	44	34	77.27	108	55	50.93
1983	55	39	70.91	117	50	42.74
1984	59	40	67.80	133	46	34.59
1985	69	43	62.32	145	53	36.55
1986	89	56	62.92	170	61	35.88
1987	88	58	65.91	165	60	36.36
1988	92	62	67.39	179	60	33.52
1989	118	66	55.93	189	59	31.22
1990	116	56	48.28	196	59	30.10
1991	134	66	49.25	235	68	28.94
1992	133	70	52.63	233	70	30.04
1993	124	72	58.06	215	68	31.63
1994	140	70	50.00	231	60	25.97
1995	138	74	53.62	223	67	30.04
1996	144	78	54.17	238	66	27.73
1997	143	73	51.05	233	60	25.75
1998	135	72	53.33	201	69	34.33
1999	151	75	49.67	190	74	38.95
2000	153	84	54.90	200	68	34.00
2001	172	78	45.35	211	69	32.70
2002	171	80	46.78	208	75	36.06
2003	170	85	50.00	214	67	31.31
2004	178	80	44.94	220	68	30.91
2005	172	82	47.67	200	73	36.50
2006	172	76	44.19	189	69	36.51
2007	170	78	45.88	212	73	34.43
2008	185	84	45.41	218	72	33.03
2009	189	83	43.92	208	73	35.10
2010	187	83	44.39	200	77	38.50
2011	185	88	47.57	206	75	36.41

DS 2 represents batters with 2 plate appearances per team game and pitchers with ⅓ innings pitched per team game between the years 1982-2011.

Method

Data Set

The data used for this research are official data of the Korean professional baseball league accumulated over a span of thirty years from 1982 to 2011. They have been acquired from the website of the Korean Baseball Organization (KBO) and *The Official Baseball Encyclopedia* published by the KBO. The yearbooks present individual player records for all batters and pitchers who are registered to the league. The play data for batters from the book consists of the number of followings: plate appearances; at-bats; runs; hits; doubles; triples; home-runs; total bases; bases on balls; intentional bases on balls; hits by pitch; strikeouts; sacrifice hits; sacrifice flies; stolen bases; caught stealing; grounded into double play; and errors, for all batters registered for a given year. From the yearbook data, only the data from pennant race games have been analyzed for this

[3] Sabermetrics or the study of baseball statistics is an independent scientific discipline named after the acronym SABR, which stands for the Society for American Baseball Research.

Table 2 | Performance measurement variables and their formulas

Abbr.	Terminology	Formula
AVG	Batting average	H / AB
OBP	On base percentage	(H+BB+HBP) / (AB+BB+HBP+SF)
SLG	Slugging percentage	[(1B)+(2X2B)+(3X3B)+(4XHR)] / AB
ERA	Earned run average	9 X ER / IP
WHIP	Walks plus hits per inning pitched	(BB+H) / IP
K/9	Strike out per 9 innings pitched	9 X (K / IP)
E/G		Errors / Games
ER/R		Earned run / Runs scored
DP/G		Double play / Games

H = Hit; AB = At bat; BB = Base on balls; HBP = Hit by pitch; SF = Sacrifice fly;
1B = Single hit; 2B = Double hit; 3B = Triple hit; HR = Home run;
ER = Earned run; IP = Inning pitched; K = Strikeout

study, while records from post season and all-star games are excluded. Batters without any computable play data except for game appearance have been excluded from the data pool. Figure 1A shows the ratio of the total number of batters to pitchers selected for the data pool by year as well as the number of regular season games per team for the last 30 years. The actual numbers are presented in Table S1. The yearbook records have been digitized and cross-checked by the co-authors of this study to ensure accuracy. Discrepancies between the KBO website data and yearbook data have been accounted for and corrected through correspondence with the KBO.

The data pool had to be further skimmed through for the purpose of this study. Since the very question addressed in this study, the extinction of 0.400 batting average, is applicable to only those who are eligible for batting titles, players eligible for batting (3.1 plate appearances per team game) and pitching (an inning pitched per team game) titles are selected as the Data Set 1 (DS 1). However, DS 1 covers only a fraction of the entire player pool (18.72 percent of all batter records and 16.52 percent of all pitcher records: See Table S2), which is considered insufficient to represent the KPB as an evolving system. In particular, eligibility for pitching titles exclude from the data set relief pitchers who play an increasingly major role in games but have relatively less innings pitched than starters. Results of analysis for the DS 1 is only presented in the supplement and used as a supportive evidence of our findings.

As a way of extending data set to better represent the KPB as a system, a new set of criteria has been established: batters with two plate appearances per team game and pitchers with a total of 1/3 innings pitched per team game have been computed as the Data Set 2. Two plate appearances per team game is the same criterion used by Gould and players selected using this represent 51.15 percent of all player records from the data pool. The percentage of DS 2 batters for any given year stays relatively stable throughout the history of KPB. The new criterion for pitchers, 1/3 innings pitched per team game, returned a rather interesting variation of representation by year—77 percent in the first year and a stable decline onward to reach 43 percent of the total pitchers in 2011. This means that for some early years, batter and pitcher representation in the DS 2 had to be asymmetrical. This can be justified by the fact that the average number of innings pitched by a pitcher has grown considerably smaller in more recent years, presumably because of clearer role demarcation between pitcher groups. Unlike the varying percentage by year, the number of pitchers selected using this criterion stays stable throughout the history of the KPB. Table 1 shows the number of players selected using each of the criteria and their ratio to the entire player pool.

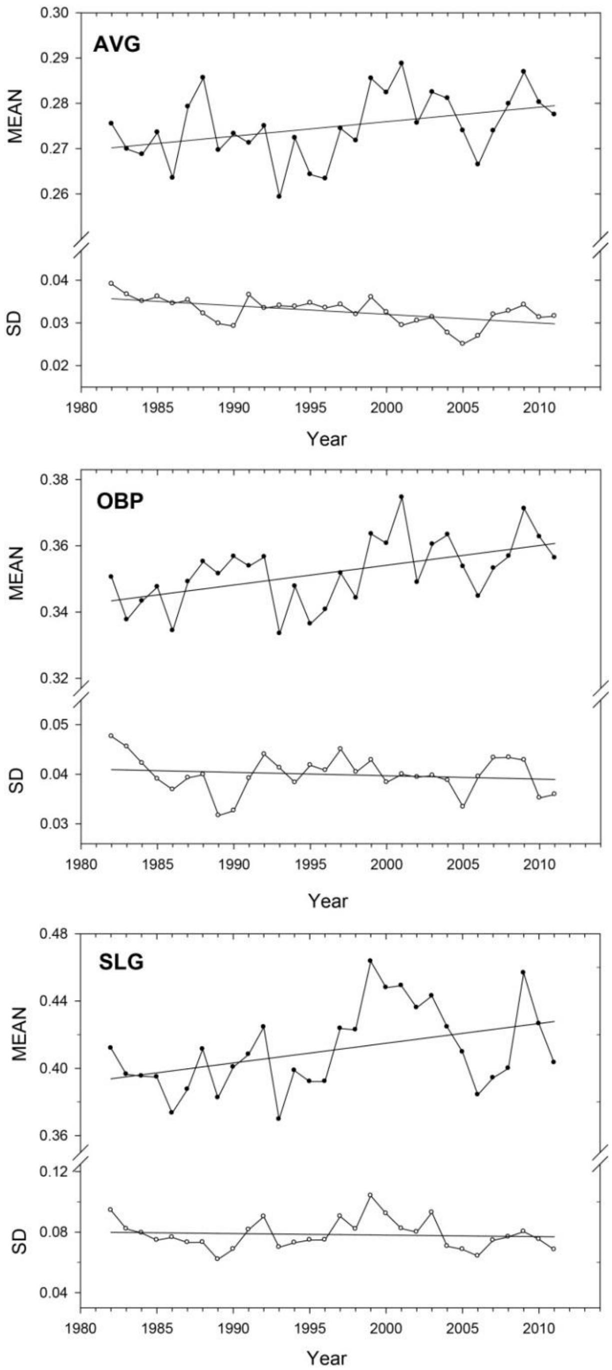

Figure 2 | Mean and standard deviation with regression line for AVG, OBP and SLG: Year 1982-2011

Table 3 | Statistical analysis of batter performance

Variable	Coefficient	Constant	Adjusted R^2
AVG_{mean}	0.0003^{*}	-0.3386	0.092
OBP_{mean}	0.0006^{***}	-0.7753^{*}	0.201
SLG_{mean}	0.0011^{**}	-1.7841^{*}	0.136
AVG_{SD}	-0.0002^{***}	0.4333^{***}	0.301
OBP_{SD}	-0.0001	0.1747	-0.009
SLG_{SD}	-0.0001	0.285	-0.026

$^{*}p<0.1$ $^{**}p<0.05$ $^{***}p<0.01$

Variables

To measure batting and pitching performance of players, a set of variables have been selected. For batters, variables selected are seasonal batting average (AVG), on-base percentage (OBP) and slugging percentage (SLG), which are a common set of offense evaluative measurements. For pitchers, earned run average (ERA), walks and hits per innings pitched (WHIP) and strikeouts per nine innings (K/9) have been selected. Formulas of the variables are provided in Table 2. For each of the variables, the annual mean and standard deviation has been computed over 30 years. Then a regression analysis has been performed for all mean and SD of variables to see whether player performance shows any distinctive trend of enhancement or decline. To determine dependence between variables, Spearman's rank correlation test was conducted for all possible combinations of batting and pitching variables. In addition to the above, fielding performance of players have been evaluated using three variables: errors per game (E/G), earned runs per runs allowed percentage (ER/R) and DP/G (double plays per game). Fielding percentage is probably a more commonly used, though often considered limited, measure of fielding effectiveness, which is calculated by dividing the total number of putouts and assists by the total number of putouts, assists and errors. Since the data from the KBO yearbooks did not provide fielder data, fielding percentage could not be calculated. Formulas for the fielding performance variables used in this study are also presented in Table 2.

Results

Figure 2 shows the results of batter data analysis, or the mean and SD of AVG, OBP and SLG data between 1982 and 2011. Table 1 displays the number and ratio of batters to pitchers in the DS 2 to the entire data pool for 30 years. The mean of annual AVG in the KPB has increased gradually for the last 30 years from around 0.27 to 0.28, or 0.0003 per year ($p < 0.1$). The mean OBP also shows an increase and the degree of increase is almost twice as steeper as that of AVG, with the annual growth standing at 0.0006 ($p < 0.01$). The SD of AVG demonstrates a discernible trend of steady decline with the coefficient standing at -0.0002 ($p < 0.01$). The decline of SD is statistically significant for all batting performance variables. The results of statistical analysis of batter performance are presented in Table 3.

Figure 3 shows changes in the statistical indexes of pitching between 1982~2011. The mean ERA has escalated, albeit moderately, for 30 years from slightly above 3.5 to well over 4.0. The mean of WHIP has fluctuated with no special tendency found, and its SD remained stable throughout 30 years. On the other hand, the mean K/9 exhibits a substantial increase, which is much sharper and more distinctive than those of ERA and WHIP. Along with the mean, the SD of K/9 shows an ascending trend as well. The results of statistical analysis of pitcher performance are presented in Table 4.

The results of Spearman's rank correlation analysis to determine the correlations between performance variables of batters and pitchers (n=30) are presented in Table 5. Except for K/9 and WHIP, all the other variable pairs are significantly correlated with each other. The same analysis has been performed for the Data Set 1, the results of which are presented in Table S3. The correlation analysis results between variables for the DS 1 appears to be similar to the analysis of DS 2.

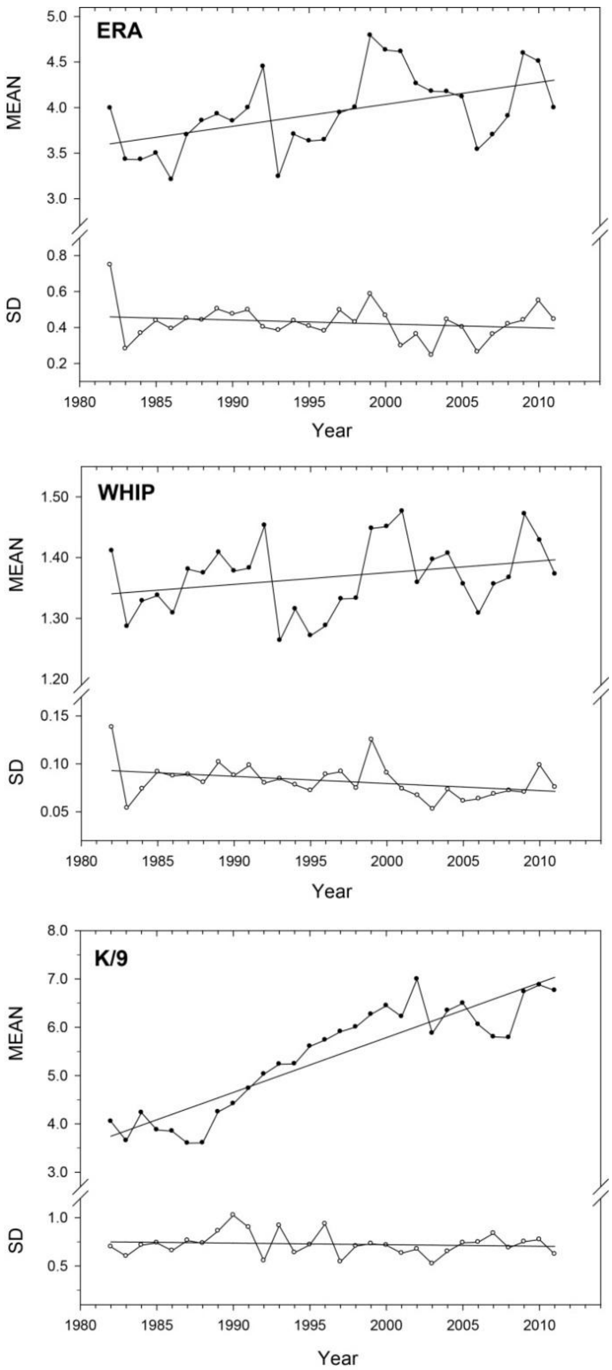

Figure 3 | Mean and standard deviation with regression line for ERA, WHIP and K/9: Year 1982-2011

Table 4 | Statistical analysis of pitcher performance

Variable	Coefficient	Constant	Adjusted R^2
ERA_{mean}	0.0241^{***}	-44.1516^{***}	0.225
$WHIP_{mean}$	0.0019	-2.4759	0.049
$K/9_{mean}$	0.1134^{***}	-221.0155^{***}	0.819
ERA_{SD}	-0.0022	4.7828	0.004
$WHIP_{SD}$	-0.0007^{*}	1.5658^{**}	0.095
$K/9_{SD}$	-0.0016	3.9224	-0.021

$^{*}p<0.1$ $^{**}p<0.05$ $^{***}p<0.01$

Table 5 | Spearman's rank correlation analysis: Batting and pitching performance measurement variables

Variable		Batter		Pitcher		
		OBP	SLG	ERA	WHIP	K/9
Batter	AVG	0.850	0.797	0.743	0.791	0.360
		<0.0001	<0.0001	<0.0001	<0.0001	0.0512
	OBP		0.741	0.787	0.882	0.446
			<0.0001	<0.0001	<0.0001	0.0134
	SLG			0.910	0.737	0.580
				<0.0001	<0.0001	0.0008
Pitcher	ERA				0.825	0.684
					<0.0001	<0.0001
	WHIP					0.290
						0.1202

In Figure 4 presented are the movements of three fielding performance variables for the last 30 years. All three variables – E/G, ER/R and DP/G – display a definitive trend of improvement. The number of E/G has been plummeted, especially for the few early years. The ratio of ER to R has risen, which means that runs allowed by error are much less in number in recent years. The average number of double plays per game has also grown gradually, with some fluctuations on the way.

Discussions

When we look at the results of batting performance analysis of the DS 2, the mean of AVG displays a slight upward movement along with the mean of OBP and SLG. That all batting performance indexes are exhibiting a steady increase indicates that batter performance in the KPB has been enhanced at all fronts, or at the very least not diminished. The SD of AVG is showing a regular decline overall with some minor fluctuations, which is consistent with Gould's findings. This is an indication of narrowing performance gaps or improving competitive balance among batters. The increase of the mean and decline of the SD appear consistent for the DS 1 analysis, or the batters qualified for batting titles (Figure S1). Overall, these results indicate stabilization and enhancement of batting performance in the Korean baseball system for the past 30 years.

All of these findings fall along the same line as Gould's theory. Since the onset of a professional baseball league in Korea, batting average of 0.400 has disappeared after a single player. In the MLB, the 0.400 hitters were relatively common in the early decades, whereas they went virtually extinct in modern baseball. Baseball critics and Sabermetrics researchers have long tried to explain the mystery by finding the cause of deterioration of batting performance in the absolute terms of batter abilities or relative to enhanced pitching performance. However, our findings show that traditional explanations are not applicable to the KPB. Batter performance in the KPB has not, by any measure, declined for the last 30 years, while the performance gaps have narrowed down steadily. We can conclude that the disappearance of 0.400 hitters is not attributable to waning batter abilities but to better competitive balance among batters or overall stabilization of batting as a system.

Pitcher performance measurements, on the other hand, returned mixed results: the mean of ERA and WHIP is increasing, albeit marginally, which may indicate that the pitcher performance in the KPB has been waning for the last 30 years. Given that the ERA and WHIP has a close correlation with all batting performance variables, decline of the mean ERA and WHIP over time is not unexpected, since batting performance is strengthening. We should note, however, that the K/9 is showing a strong increase, and the result is consistent for the DS 1 (See Figure S2. This leaves some room for interpretation, which will be discussed in detail in the next paragraph. The increase of K/9 is an indication that pitchers have grown better at striking batters out, although it is still a bit of a stretch to say that pitching performance in the KPB has enhanced. We can draw from this that enhancement of pitcher performance was not the cause of the disappearance of 0.400 hitters in the KPB. The SD tends to decline for all variables, but the trends are not statistically significant except for that of WHIP. Overall, the mixed results suggest that pitcher performance in the KPB can be still in the process of stabilization. Part of the reason can be the obvious increase of the number of pitchers and clearer role division between feature groups as indicated in Figure 5. The ratio of pitchers with pitching title eligibility to the entire pitcher pool is decreasing exponentially. The runs allowed per game (Figure 6) is increasing, which could mean the overall balance of pitching is still shaky. Further analysis of pitcher performance in the KPB can be conducted when more data are accumulated with passing time.

Regarding the sharp increase of the mean K/9, the first and the most obvious explanation is that pitcher performance, although its balance with batting has been tilting toward the other side, has not waned in absolute terms at least at some fronts after all. Since the correlation with batting performance variables is the weakest for K/9 among all three pitcher performance variables, it can be considered relatively independent as a measure of pitcher performance. A previous study concludes that K/9 is a better evaluative measure of a pitcher's ability (Albert, 2010) than ERA

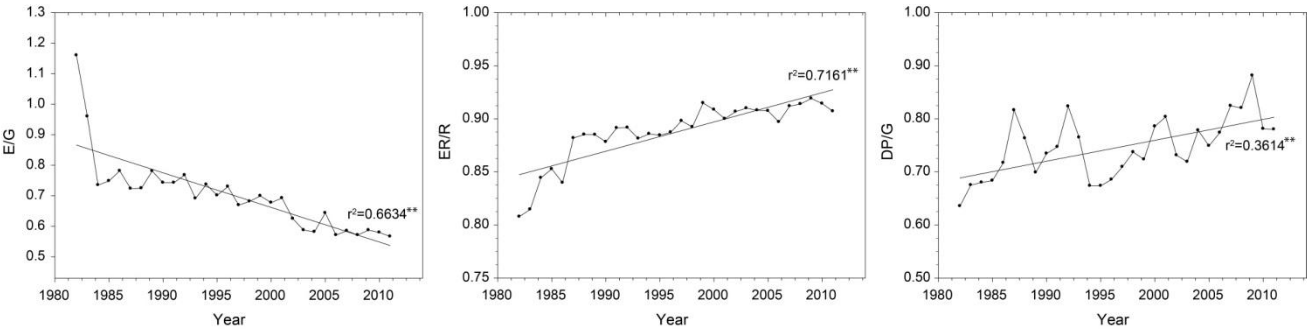

Figure 4 | Mean with regression line for E/G, ER/R and DP/G: Year 1982-2011

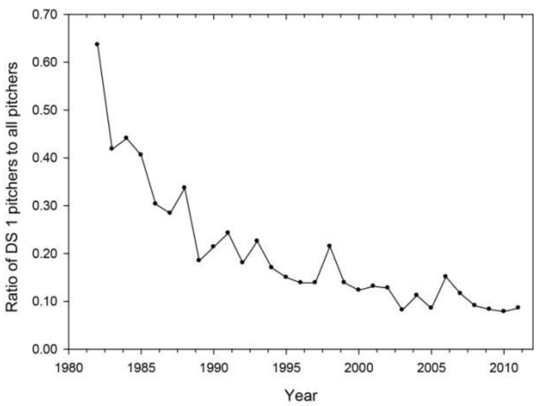

Figure 5 | Ratio of pitchers qualified for pitching titles to all pitchers: Year 1982-2011

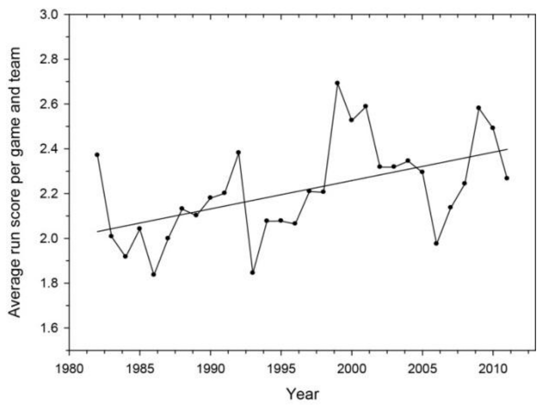

Figure 6 | Average runs scored per game per team: Year 1982-2011

and WHIP. An alternative explanation is that the obvious increase may be the result of change in the batting style. Batters aiming for "biggies," or who are more commonly described as sluggers, tend to be struck out more frequently due to the use of big, flat-armed swing. When we go back to the batter data analysis, the mean of SLG is increasing, which is supportive of this second explanation. Also for the K/9, the correlation with SLG is relatively stronger than that with other batting performance variables, which may indicate that the increase of K/9 is the direct result of more attempts at slugging at the batters' end. However, the increase of the mean K/9 can be also accounted for by introduction of a variety of breaking balls. To further support the correlation between K/9 and slugging, we need to look into the strikeout data in more detail. The strikeout data used for this study made no distinction between strikeouts-looking and swing-strikeouts. By analyzing play-by-play data, the correlation between K/9 and slugging in the KPB will be better validated and we leave this for a future study.

Fielding performance variables for the last 30 years exhibits a surprising tendency of sharp increase. Gould also observed mean fielding average in the MLB to support his arguments for performance enhancement over time, saying if baseball has improved, he should be able to "note a decelerating rise in fielding averages" (Gould, 1996), and indeed he did. We can observe the same trend in the improvement of the mean of E/G and ER/R, which is much more prominent for the early years (1982-1990) and then balances out to gradual stabilization in more recent years. It is also noteworthy that the ratio of ER to R has been dramatically increasing, or less and less runs are allowed due to errors. This indicates that the influence of errors during game play is waning. Overall, the fielding performance variables can be described as the most telling measure of improvement and stabilization of the KPB over time.

The methodology we employed to reach the conclusion above has certain analytical limitations: Observing the annual mean and the standard deviation of variables over time is a relatively simple approach to statistical evaluation of performance measurement, but the approach does not take into account discrepancy of opportunities, i.e. the number of at-bats or innings pitched for a player. In statistical analysis, it is questionable whether the simple mean of batting averages or ERA is still valid without consideration of distribution. To resolve this analytical problem, a weighted average or weighted standard deviation of variables can be adopted in consideration of the distribution of the number of opportunities for each player.

Although analysis of the KPB data returned results largely in accord with Gould's findings on the MLB, there are some minor differences between the results we presented above and Gould's. As we discussed previously, the decrease of SD is relatively indistinctive and irregular in the KPB. This may be attributable to the fact that baseball as a game was much more established in 1982 than in the 19th century. To cite an example, there have been no major changes to game rules in the KPB that could have significantly impacted the batting-pitching balance, such as the distance of mound from the home plate[4]. Also in the MLB, aggressive global scouting efforts and huge payoff entice the best of the bests, whereas in the KPB, the source of player pool is relatively limited. Starting in 1994, the top-tier baseball talents in Korea made it to the leagues outside the country, a phenomenon unknown in the MLB. All of these suggest interesting avenues for future studies. This study can be viewed as to provide a first step towards understanding the changes that have taken place in the KPB for the last 30 years. The next step would be to explore the unique characteristics of the KPB to gain better understanding of its evolution in relation to other leagues.

Parameters that may have affected individual player performance but are difficult to quantify have been excluded from this study, including the size of stadiums, time of the day when a game takes place (daytime or nighttime) or the factor of double header games. All the data from the main analysis have been collectively analyzed and did not take into account characteristics of individual players.

The variables used for this study can be described as the most common set of evaluative measurements of offense and defense performance in baseball, but they could be extended to include studies of other evaluative measurements of player performance. Studies of factors and parameters mentioned above are also promising avenues of future studies that can take advantage of the data verified by this study. We hope that this study will lay groundwork for vigorous statistical analysis of the Korean professional baseball and eventually contribute to the success of the league.

[4]In the MLB, the pitching mound was moved backwards at the end of the 19th century, after which the mean batting average soared and stayed high until foul-strikeout was adopted.

Acknowledgement

We would like to express our sincere gratitude to the Korea Baseball Organization and SPORTS2i for the base data and other generous support they provided us with.

References

Albert J (2010) Dolciani Mathematical Expositions: Sabermetrics: The Past, the Present, and the Future. *Mathematics and sports* 3-15.

Bernstein PL (1998) Where, Oh where are the .400 hitters of yesteryear? *Financial Analysts Journal* 54: 6-14.

Berry DJ, Brook SL, Frick B, Fenn AJ, Vincent-Mayoral R (2005) The short supply of tall people: Competitive imbalance and the national basketball association. *Journal of Economic Issues* 34: 1029-1041.

Chatterjee S, Hawkes J (1995) Apparent decline as sign of improvement? Or, can less be more? *Teaching Statistics* 17: 82-83.

Gould SJ (1996) *Full house: The spread of excellence from Plato to Darwin*. New York: Harmony Books.

LeonardII WM (1995) The decline of the .400 hitter: an explanation and a test. *Journal of Sport Behavior* 18: 226-236.

Miceli TJ (2005) Minimum quality standards in baseball and the paradoxical disappearance of the .400 hitter. *Economics Working Papers*: 200515.

Radicchi F (2011) Who is the best player ever? A complex network analysis of the history of professional tennis. *Plos One* 6: E17249.

Schall T, Smith G (2000) Do baseball players regress toward the mean? *American Statistician* 54: 231-235.

Schmidt MB, Berri DJ (2005) Concentration of playing talent: Evolution in Major League Baseball. *Journal of Sports Economics* 6: 412-419.

Williams T, Underwood J (1986) *The Science of Hitting*. New York: Simon and Schuster.

Supplementary Information

Table S1 | Number of teams, games, players and spectators of the Korea Professional Baseball league for the past 30 years

Year	Number of teams	Number of team games	Number of batters	Number of pitchers	Number of spectators
1982	6	80	109	44	1,438,768
1983	6	100	117	55	2,256,121
1984	6	100	133	59	1,664,720
1985	6	110	145	69	1,688,365
1986	7	108	172	89	2,141,112
1987	7	108	167	88	2,019,675
1988	7	108	180	92	1,932,145
1989	7	120	191	119	2,883,669
1990	7	120	199	117	3,189,488
1991	8	126	239	136	3,825,409
1992	8	132	240	133	3,912,092
1993	8	132	218	124	4,437,149
1994	8	132	235	141	4,194,428
1995	8	132	228	140	5,406,374
1996	8	132	247	144	4,498,082
1997	8	132	239	144	3,902,966
1998	8	132	211	135	2,639,119
1999	8	132	203	151	3,220,624
2000	8	133	208	154	2,507,549
2001	8	133	224	175	2,991,064
2002	8	133	225	172	2,394,570
2003	8	133	221	171	2,722,801
2004	8	133	227	178	2,331,978
2005	8	126	210	175	3,387,843
2006	8	126	197	172	3,040,254
2007	8	126	224	172	4,104,429
2008	8	126	276	187	5,256,332
2009	8	133	243	192	5,925,285
2010	8	133	277	191	5,928,626
2011	8	133	217	186	6,809,965

Table S2 | Number and percentage of players in data set 1 (DS1)

Year	Number of pitchers			Number of batters		
	All	DS 1	DS 1 / All (%)	All	DS 1	DS 1 / All (%)
1982	44	28	63.64	153	37	34.26
1983	55	23	41.82	172	33	28.21
1984	59	26	44.07	192	30	22.56
1985	69	28	40.58	214	31	21.38
1986	89	27	30.34	261	35	20.59
1987	88	25	28.41	255	35	21.21
1988	92	31	33.70	272	33	18.44
1989	118	22	18.49	310	34	17.99
1990	116	25	21.37	316	36	18.37
1991	134	33	24.26	375	35	14.89
1992	133	24	18.05	373	33	14.16
1993	124	28	22.58	342	29	13.49
1994	140	24	17.02	376	22	9.52
1995	138	21	15.00	368	25	11.21
1996	144	20	13.89	391	31	13.03
1997	143	20	13.89	383	33	14.16
1998	135	29	21.48	346	42	20.90
1999	151	21	13.91	354	45	23.68
2000	153	19	12.34	362	42	21.00
2001	172	23	13.14	399	47	22.27
2002	171	22	12.79	397	39	18.75
2003	170	14	8.19	392	40	18.69
2004	178	20	11.24	405	48	21.82
2005	172	15	8.57	385	42	21.00
2006	172	26	15.12	369	38	20.11
2007	170	20	11.63	396	44	20.75
2008	185	17	9.09	463	39	17.89
2009	189	16	8.33	435	43	20.67
2010	187	15	7.85	468	45	22.50
2011	185	16	8.60	403	36	17.48

DS 1 represents batters with 3.1 plate appearances per team game and pitchers with 1 inning pitched per team game between the years 1982-2011.

Table S3 | Spearman's rank correlation analysis: Batting and pitching performance measurement variables for data set 1 (DS1)

Variable		Batter		Pitcher		
		OBP	SLG	ERA	WHIP	K/9
Batter	AVG	0.855 <0.0001	0.704 <0.0001	0.663 <0.0001	0.761 <0.0001	0.3386 0.035
	OBP		0.710 <0.0001	0.754 <0.0001	0.749 <0.0001	0.539 0.002
	SLG			0.769 <0.0001	0.615 <0.0001	0.704 <0.0001
Pitcher	ERA				0.839 <0.0001	0.670 <0.0001
	WHIP					0.360 0.051

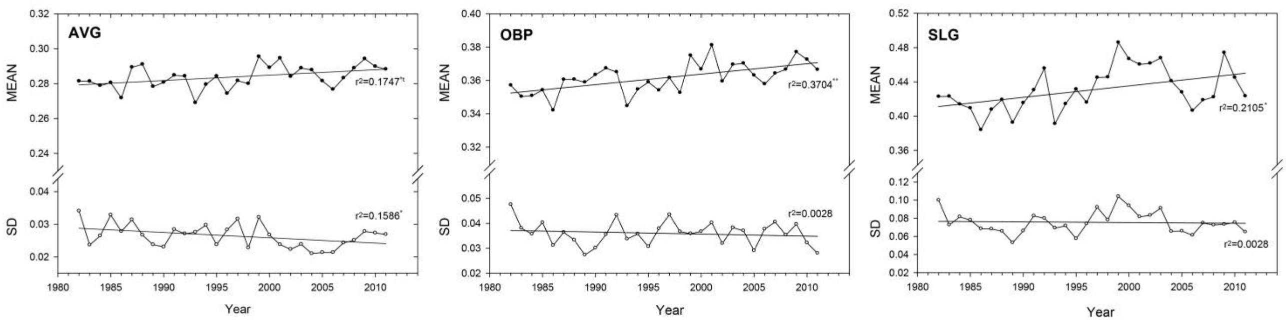

Figure S1 | Mean and standard deviation with regression line for AVG, OBP and SLG of data set 1: Year 1982-2011

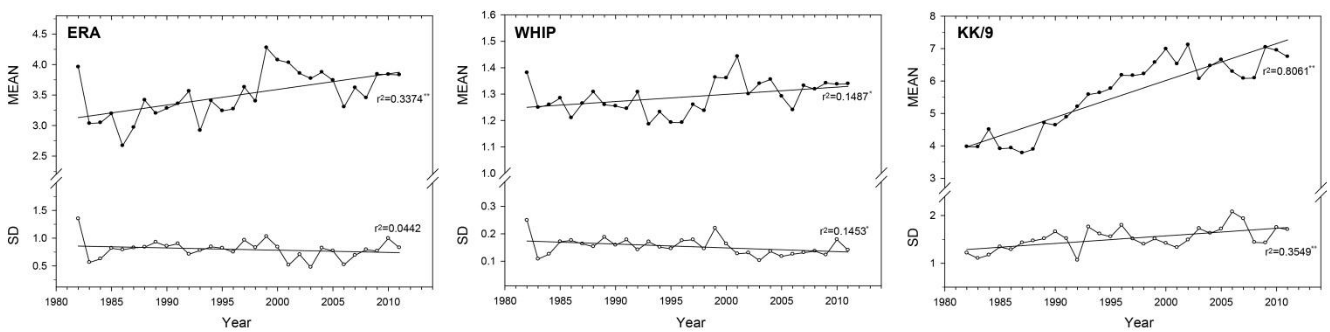

Figure S2 | Mean and standard deviation with regression line for ERA, WHIP and K/9 of data set 1: Year 1982-2011