# WORKING LIFE EXPECTANCY OF MAJOR LEAGUE PITCHERS AND FORECASTING THE NUMBER OF THEM: TASKS MADE EASY BY USING THE COHORT CHANGE RATIO METHOD

David A. Swanson (Corresponding Author)
University of California Riverside
dswanson@ucr.edu
951-534-6336

Jack Baker
HealthFitness Corporation
quantmod@gmail.com
505-250-9619

Jeff Tayman
University of California San Diego
jtayman@ucsd.edu
858-581-6710

Lucky M. Tedrow
Western Washington University
Lucky.Tedrow@wwu.edu
360-220-9072

**Abstract**
The lack of information on major league baseball career length is surprising given the detailed historical data on baseball players and the large number of working life tables for various occupations. This deficit may be due at least in part to an assumption that the process needed to construct MLB working life tables is such a demanding task that it discourages many from the attempt. If so, we believe this paper shows that such an assumption needs to be re-examined by showing how easy it is to use the cohort change ratio method to estimate a working life table of major league baseball pitchers for the period 1980-81. We also generate a short term forecast for 1982 using the same cohort change ratios. We conclude that this paper has demonstrated that both reasonable estimates of MLB working life expectancies and forecasts can be easily constructed by using this method. The ease of using it may serve to fill in gaps concerning both baseball working life expectancies and future player numbers in the major leagues, an outcome that may serve to help address similar knowledge gaps that exist in other sports, both at the professional and amateur levels.

### *Background*

Life tables have been constructed for the working life expectancy of major league baseball (MLB) players (Witnauer et al. 2007) and regression models have been constructed by Hardy et al. (2017) to assess the determinants of career length among major league pitchers playing between 1989 and 1992. In addition, other studies have looked at general life expectancy among MLB players (Abel and Kruger 2005, Saint Onge et al. 2008) as well as elite athletes in other sports (Sarna and Kaprio 1994, Sarna et al. 1993, and Schnohr 1971). However, it is the case that working life tables for MLB pitchers have not been constructed. As such, little is known about their career prospects from this empirical perspective. Witnauer et al. (2007) found this lack of information surprising given the detailed historical data on baseball players and the large number of working life tables for various occupations. We speculate that this deficit may be due at least in part to an assumption that the process needed to construct MLB working life tables is such a demanding task that it discourages many from the attempt. This paper is aimed at showing that such an assumption needs to be re-examined.

Thus, in this paper, we provide information about the career prospects of MLB pitchers by using a life table perspective and in so doing, show how easily this can be done by using the cohort change ratio (CCR) method. We illustrate this process by constructing a working life table of MLB pitchers for the period 1980-81 using data from the eighth edition of *Total Baseball*. Using the same data, we also show how easy it is to generate a short term forecast using the CCR method in conjunction with the working life table. We illustrate this latter use with a one year forecast of major league pitchers by consecutive years in MLB, an application that like the working life table for MLB pitchers, appears to be novel.

In the paper that follows, we first describe CCRs in general, describe how they can be used to generate population projections and then describe how CCRs can be used to construct life tables. Following these methodological discussions, we first construct and discuss a working life table for MLB pitchers and then construct and discuss a forecast of them. We conclude the paper with remarks on the results and suggestions for future research and applications.

### *Cohort Change Ratios*

As shown by Baker et al. (2017) CCRs have a wide range of applications. When migration is negligible, they can be used to construct life tables and calculate life expectancy (Baker et al., 2017: 165-171). These same CCRs also can be used to generate forecasts (Baker et al., 2017: 45-58).[1]

A cohort change ratio (CCR) is typically computed from age-related data in the two most recent censuses (Baker et al. 2017: 2):

$${}_{n}\mathrm{CCR}_{x,\,t} = {}_{n}\mathrm{P}_{x,t} \,/\, {}_{n}\mathrm{P}_{x\text{-}k,\,t\text{-}k.} \qquad [1]$$

where,

${}_{n}P_{x,\,t}$ is the population aged $x$ at the most recent census ($t$),

${}_{n}P_{x-k,\,t-k}$ is the population aged $x$-$k$ at the 2nd most recent census ($t$-$k$), and

$k$ is the number of years between the most recent census at time $t$ and the one preceding it at time $t$-$k$.

As implied by Eq. [1], a cohort change ratio is not typically computed for a single cohort, but for all of the cohorts found in two successive census counts.

Given the nature of the CCR, 10-14 is the youngest five-year age group for which CCRs as defined in Eq. [1] can be made if there are 10 years between censuses. To analyze age groups younger than ten in a given application, a Child-Adult Ratio (CAR) can be used. This ratio, computed separately for ages 0-4 and ages 5-9, relates young children to adults in the age groups most likely to be their parents (Baker et al. 2017: 3, Smith et al. 2013: 178).

The open-ended age group uses the same approach found in life table construction (Baker et al. 2017: 3), and its CCR differs slightly from those for the age groups beyond age 10 up to the oldest open-ended age group. If for example the final closed age group is aged 70-74, with persons aged 75+ as the terminal open-ended age group, then calculation for the $CCR_{x+,t}$ requires the summation of the three oldest age groups to get the population age 65+ at time *t-k*:

$${}_{\infty}CCR_{75,t} = {}_{\infty}P_{75,t} / {}_{\infty}P_{65,t-k}. \qquad [2]$$

### *Using CCRs to estimate Life Expectancy*

Baker et al. (2017: 165) show that when migration is negligible the CCR can be interpreted as a census survival ratio, which means the expectation of life at age x can be computed as:

$$e_x = (T_x / l_{(n/2)}) / (l_x / l_{(n/2)}) = T_x / l_x \qquad [3]$$

where,
$x$ is age,
$n$ is the width of the age groups (up to, but not including the terminal, open-ended age group),
$e_x$ is the life expectancy (average years remaining) at age $x$,
$T_x$ is the total person years remaining to persons age $x$,
$l_x$ is the number of persons reaching age $x$,
$l_{(n/2)}$ = persons aged $x$ to $x+n$ are assumed to be concentrated at the mid-point of the age group, and

$$l_{(x+n/2)} / l_{(x-n/2)} = P2_{(x,n)} / P1_{(x-n,n)} \qquad [4]$$

where,
$P2_{(x,n)}$ is the number of persons counted in the second census in age group $x$ to $x+n$, and
$P1_{(x-n,n)}$ is the number of persons counted in the first census in age group $x-n$ to $x$.

In general, the life-table probability of surviving from the mid-point of one age group to the next ($l_{(x+n/2)} / l_{(x-n/2)}$) is approximated by the census survival ratio ($P2_{(x,n)} / P1_{(x-n,n)}$). Continuing, the cumulative multiplication of the probabilities shown in [4] gives the conditional survival schedule ($l_x / l_{(n/2)}$). From the conditional $l_x$ values given by [4] the conditional estimates of the number of person years lived in each age group (${}_nL_x$) can be calculated as:

$${}_nL_x / l_{(n/2)} = (n / 2) \times [(l_x / l_{(n/2)} + l_{(x+n)} / l_{(n/2)}] \qquad [5]$$

where,
${}_nL_x$ is the number of person years lived in each age group.
Given a value of $T_x / l_{(n/2)}$ for some initial age x, total remaining years expected at age $x$ ($T_x$) values can be calculated as:

$$T_{(x-n)} / l_{(n/2)} = T_x / l_{(n/2)} + {}_nL_{(x-n)} / l_{(n/2)}. \qquad [6]$$

This leads us back to equation [3], so that the expectation of life at age x is:
$e_x = (T_x / l_{(n/2)}) / (l_x / l_{(n/2)}) = T_x / l_x$.
Extending this approach, we note that when the radix of a life table is equal to 1 ($l_0 = 1.00$) life expectancy at birth can be computed directly from the expression:

$$e_0 = S_0 + (S_0 \times S_1) + (S_0 \times S_1 \times S_2) +,..., + (S_0 \times S_1 \times S_2,..., \times S_x) \qquad [7]$$

where,
$e_0$ is the life expectancy at birth,
$S_0$ is the survivorship from $t=0$ (e.g., birth) to $t=1$(e.g., age 1),
$S_1$ is the survivorship from $t = 1$ (e.g., age 1) to $t = 2$(e.g., age 2), and so on through $S_x$ , and
$S_x$ is ${}_1L_x$ / ${}_1L_{(x-n)}$.
Equation [7] is set up for single year age groups. However, we can generalize it to other age groups: ${}_nS_x = {}_nL_x / {}_nL_{(x-n)}$, so that:

$$e_0 = {}_nS_0 + ({}_nS_0 \times {}_nS_1) + ({}_nS_0 \times {}_nS_1 \times {}_nS_2) +,...,+ ({}_nS_0 \times {}_nS_1 \times {}_nS_2,..., \times {}_nS_x) \quad [7.a]$$

As equations [7] and [7.a] both imply, the fundamental life table function is inherent in our method; that is, via the ${}_nS_x$ values, we have ${}_nq_x$ values. In summary, our approach is the result of combining either equation [7] or [7.a] for computing life expectancy with equation [1] to estimate $e_x$. Broadly speaking, the method can be applied to any population subject to renewal through a single increment (entry into the major leagues) and extinction through a single decrement (exit from the major leagues), where there are at least two successive counts that provide the population by some measure of time (consecutive years in the major leagues).

### *Working Life Table*

Although baseball is the "national sport," only a handful of studies have examined its career prospects in terms of working life. Using data from 1902 to 1993, Witnauer et al. (2007) found that non-pitching rookie position players can expect to play 5.6 years. Abel and Kruger (2005) examined the overall life expectancy of MLB players by position for the period 1909-1919 and found that pitchers could expected to live to age 67.7, which exceeded the life expectancy of males (63.4) for this same period. While they did not construct a life table, Hardy et al. (2017) used a regression-based approach to assess the determinants of career length among major league pitchers playing between 1989 and 1992 and found that mean career length was 10.97 years. Truncated careers can be attributed to several factors, including injury, poor performance, and scandals (Gutman 1992, Hardy et al. 2017).

Using the method described in the preceding section and 1980 and 1981 data found in the eighth edition of *Total Baseball* (Thorn 2004), Table 1 shows the working life expectancy of major league pitchers who entered the major leagues in 1980.[2] This table shows the number of

(Table 1 about here)

pitchers who "survive" ($l_x$) by season (year $x$, where $x = 0$ to $10+$) starting with the 296 pitchers who first entered MLB in 1980 ($l_0$). Of these 296 pitchers, 247 completed the first year of play and only 109 (44%) completed five consecutive years. The ($S_x$) column shows the proportion surviving through each year, which can be interpreted as the probability of making it through the entire season. The probability of making it through the first year is ($S_0$) = 0.83333 (247 / 296). Corresponding to $S_0$, the number of "exits" ($d_0$) in the first year among the 296 initial pitchers is 49, which corresponds the probability of "exiting" in the first year is ($q_0$) = 0.166667 (49 /296). At the start of each year, the expected remaining consecutive years of pitching is provided. For the initial 296 pitchers, their expected working life is 3.990 years. At the completion of the first year, those still pitching can expect to do so for 3.157 more years. Those who complete five consecutive years can expect to pitch for 1.172 more years.

This life table shows that the expected working life of MPB pitchers is short and becomes shorter at the completion of each year. Witnauer et al. (2007) found that, in general, a (non-pitching) player who enters MLB can expect to play 5.6 years. Moreover, they found that those who made it through three years could expect to play an additional six years. MLB pitchers, however, who complete three years can expect to play for only 2.114 years more.

### *A Forecast of Major League Pitchers*

In our illustration of the application of the CCR forecasting method, we forecast major league pitchers according to the number of consecutive years they have been in the major leagues. As was the case for the life table, we are not aware of any forecasts of major league pitchers. Years in the majors are measured by consecutive integers ranging from 0 to 10+ years, making them analogous to single-years of age. As such, we can construct CCRs using two consecutive annual time points for each amount of time in the majors.

In this illustration, we use historical information for 1980 and 1981 to develop a one-year (1982) forecast of major league pitchers by consecutive years in the majors (see Table 2). These years were not affected by changes in the number of major league teams (there were 26 in each of the three years) and the three-year set allows us to compare the 1982 forecast to the recorded 1982 numbers to get an idea of the accuracy of using the method for this purpose. It also is the case that the current "luxury tax" found in major league baseball does not affect the forecast because the tax was not introduced until 1996 (Dietl et al. 2010). This also means that a forecast constructed using data subsequent to 1995 and prior to, say, 2015 would have a mixture of effects representing the pre-tax era and the luxury tax era. It would only by using data subsequent to 2015 that a forecast would likely have data that only reflect the luxury tax era.

(Table 2 about here)

To assemble the data found in this illustration, we used a hardcopy of the eighth edition of *Total Baseball* (Thorn 2004), which means that identifying the correct data and transcribing it (into Excel) was subject to a high level of potential error, given what we wanted to record. To minimize transcription error we developed and used the following protocol. First, we identified all of the pitchers who played in any one of the years of interest, 1980, 1981, and 1982 in the hardcopy edition. We next weeded out those who played in those years but did not play consecutively in prior years (In his 27 year career, Nolan Ryan, for example, was in the majors in each of the three years of interest, but because he initially is listed in the majors in 1966 and then not listed again until 1968, he is not included in our data. Following these two steps, we then did three separate counts for each of the three years of interest. Between the first and second count we weeded out additional errors (e.g., in the first pass we missed Scott Brown who is listed only in 1981 and erroneously included Joey McLaughlin. In the second pass, we realized that Scott Brown had been missed while Joey McLaughlin should not have been included because he first was listed in 1977 and then in 1979, 1980, 1981, and 1982, but not in 1978. As such, McLaughlin was deleted because he did not have consecutive years in the majors between the year he was first listed and any of the three years of interest, 1980, 1981, and 1982. By the third count, the numbers matched up with the second count and we were satisfied that the data we had assembled were sufficiently accurate to use. As an additional test, we counted the number of pitchers by year found in a third count that was separate from the count by number of consecutive years played and then compared these counts with the sum of the number by year and consecutive years played as found in the third set of counts. For 1980, the counts matched at 308, as did the count for 1981 (296) and the count for 1982 (293).

(Table 2 about here)

A traditional CCR cannot be computed for zero years in the majors, so we averaged the ratio of pitchers with zero consecutive years in the league to pitchers with one and two consecutive years in the league for 1980 and 1981. CCRs for the other years spent in the league are analogous to traditional CCRs. For example, the CCR for 4 consecutive years in the majors is computed by dividing number of pitchers in the league for 4 consecutive years in 1981 by the number of pitchers in the majors for 3 consecutive years in 1980 (0.93939 = 31 / 33). The CCR for the open-ended category (10+ years) is the number of pitchers with 10 or more consecutive years in the majors in 1981 divided by the number of pitchers with 9 or more consecutive years in the in 1980 (0.78571 = 44 / (12 + 44)).

The 1982 forecast for the all pitchers is 11 lower than the actual count, for an error of -3.8%. There is a wide range of errors for the individual consecutive years in the league, ranging in absolute terms from 0.0% to 25.0%. Errors for 7 categories are less than 8.0%, with the other four categories showing double digit percentage errors. On average, the forecast has a slight downward bias with a MALPE of -1.0% and a lower degree of accuracy with a MAPE of 8.7%.

### *Concluding Remarks*

The data shown in Tables 1 and 2 indicate that there is a high level of volatility in the major league careers of pitchers, especially in the initial years. The factors causing this likely include: (1) injuries that lead to one or more missed seasons; (2) being sent down to the minors for one or more seasons to gain more experience, (one common example is that the initial listing represents a pitcher who is "called up" for a couple of games at the end of the season to "have a cup of coffee," followed by a return to the minors); and (3) either outright release or a decision to quit professional baseball made by the player. Given this volatility, this example can be viewed as a rather strenuous test of how well the CCR method can perform in subject areas where there is less stability year to year than found in large populations. These areas would include any highly competitive activity such as professional sports. In this regard, while there is substantial variation in the careers of major league such variation can be smoothed by fitting a model to the data. Given that both the life table and forecast of major league pitchers used survival rates, the Weibull model may be a good candidate (Namboodiri and Suchindran 1987).

Even without smoothing the data, the accuracy of the 1982 forecast by number of consecutive years in the majors is reasonably accurate with an overall error of only -3.8%. Given this, the CCR method may work reasonably well as a method to forecast the overall number of players in other professional sports by position, such as football (e.g., quarterbacks), basketball (e.g., point guards), hockey (e.g., goalies), and soccer (e.g., strikers). Continuing along this line of reasoning, it may be worthwhile to examine the CCR method in terms of occupational specialties in organizations (e.g., corporations, governments, the armed forces). Forgetting positions or occupational specialties, the method also may be worthwhile examining in terms of participants classified by years played in a number of activities, including tennis, golf, and NASCAR racing. Keep in mind that at the professional level, some of these sports are affected by luxury taxes, salary caps and other financial restrictions, while others are not (Dietl et al. 2010). As such, the history of the implementation of these measures may be important in terms of constructing a working life table or otherwise estimating expected career length and, especially, in terms of forecasting.

In conclusion, we observe that the aim of this paper was to demonstrate that both reasonable estimates of MLB working life expectancies and forecasts can be easily constructed by using the CCR method. If it does change assumptions about these being demanding tasks, perhaps the knowledge gaps concerning working MLB life expectancies and future player numbers can be filled in, an outcome that may serve to help address similar knowledge gaps that exist in other sports, both at the professional and amateur levels.

**Endnotes**

1. In terms of forecasting, the CCR approach falls into the simple category rather than the complex category (Green and Armstrong 2015); and while there is no evidence that shows complexity improves forecast accuracy, Green and Armstrong (2015) suggest it remains popular among: (1) researchers, because they are rewarded for publishing in highly ranked journals, which favor complexity; (2) forecasters, because complex methods can be used to provide forecasts that support decision makers' plans; and (3) clients, who may be reassured by incomprehensibility. Green and Armstrong (2015) suggest that clients who prefer accuracy should use forecasts only from simple evidence-based procedures.

2. The addition of two teams (Colorado Rockies and Florida Marlins) to the National League in 1993 may have extended slightly the career of some of the pitchers who had played consecutively since 1981. Because we used ten as the terminal, open-ended consecutive years played, there is, however, no effect on the working life table by this expansion. Similarly, there is no effect on the working life table by subsequent expansions.

**References**

Abel, E. and M. Kruger. (2005). Longevity of major league baseball players. *Research in Sports Medicine* 13: 1-5.

Baker, J., Swanson, D., Tayman, J., and Tedrow, L. (2017). *Cohort change ratios and their applications*. (2017). Springer: Dordrecht, The Netherlands.

Dietl, H., M. Lang, and S. Werner. (2010). The effect of luxury taxes on competitive balance, club profits, and social welfare in sports leagues. *International Journal of Sport Finance* 5(1), 41-51.

Green, K. and J.S. Armstrong (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research* (68): 1678-1685.

Gutman, D. (1992). *Baseball Babylon*. New York: Penguin Books.

Hardy, R., T, Ajibewa, R. Bowman, and J. Brand. (2017) Determinants of major league baseball pitchers' Career Length. *The Journal of Arthroscopic and Related Surgery* 33, 445-449.

Namboodiri, K. and C. Suchindran. (1987). *Life table techniques and their applications*. Academic Press: San Diego, CA.

Saint Onge, J., R. Rogers, and P. Krueger. (2008). Major league baseball players' life expectancies. *Social Science Quarterly* 89(3), 817-830

Sarna, S. and J. Kaprio (1994). Life expectancy of former elite athletes. *Sports Medicine* 17, 149–151.

Sarna S.,T. Sahi, M. Koskenvuo, and J. Kaprio. (1993). Increased life expectancy of world class male athletes. *Medicine and Science in Sports and Exercise* 25, 491–495.

Schnohr, P. (1971) Longevity and causes of death in male athletic champions. *Lancet* 2, 1364–1365.

Smith, S. K., J. Tayman, J., & D. Swanson. (2013). *A practitioner's guide to state and local population projections*. Springer: Dordrecht, The Netherlands.

Thorn, J. (2004). *Total baseball: The ultimate baseball encyclopedia (8th edition).* Sport Media Publishing.

Witnauer, W., R. Rogers, and J. Saint Onge (2007). Major league baseball career length in the 20th century. *Population Research and Policy Review* 26, 371-386.

**Table 1. Major league pitcher working years life table, 1980-1981**

| Consecutive years in the majors | Probability of pitching next year $S_x$ | Probability of not pitching next year $q^x$ | Number of pitchers lx | Pitchers leaving the majors dx | Pitching years expectancy $e_x$ |
|---|---|---|---|---|---|
| 0 | 0.83333 | 0.16667 | 296 | 49 | 3.990 |
| 1 | 0.70723 | 0.29277 | 247 | 72 | 3.157 |
| 2 | 0.76923 | 0.23077 | 175 | 40 | 2.568 |
| 3 | 0.93939 | 0.06061 | 135 | 8 | 2.114 |
| 4 | 0.85714 | 0.14286 | 127 | 18 | 1.688 |
| 5 | 0.85714 | 0.14286 | 109 | 16 | 1.172 |
| 6 | 0.90000 | 0.10000 | 93 | 9 | 0.859 |
| 7 | 0.83333 | 0.16667 | 84 | 14 | 0.578 |
| 8 | 0.81818 | 0.18182 | 70 | 13 | 0.343 |
| 9 | 0.78571 | 0.21429 | 57 | 12 | 0.151 |
| 10+ | 0.00000 | 1.00000 | 45 | 45 | 0 |

**Table 2 Forecast of pitchers by the number of consecutive years in the major leagues, 1982**

| Consecutive years in the majors | 1980 | 1981 | CCR[a] | 1982 Forecast[b] | Actual | Numeric Error | Percentage Error |
|---|---|---|---|---|---|---|---|
| 0 | 48 | 44 | 0.61884 | 40 | 48 | -8 | -16.7% |
| 1 | 41 | 40 | 0.83333 | 37 | 39 | -2 | -5.1% |
| 2 | 39 | 29 | 0.70732 | 28 | 33 | -5 | -15.2% |
| 3 | 33 | 30 | 0.76923 | 22 | 21 | 1 | 4.8% |
| 4 | 21 | 31 | 0.93939 | 28 | 28 | 0 | 0.0% |
| 5 | 21 | 18 | 0.85714 | 27 | 24 | 3 | 12.5% |
| 6 | 20 | 18 | 0.85714 | 15 | 15 | 0 | 0.0% |
| 7 | 18 | 18 | 0.90000 | 16 | 17 | -1 | -5.9% |
| 8 | 11 | 15 | 0.83333 | 15 | 12 | 3 | 25.0% |
| 9 | 12 | 9 | 0.81818 | 12 | 13 | -1 | -7.7% |
| 10+ | 44 | 44 | 0.78571 | 42 | 43 | -1 | -2.3% |
| Total | 308 | 296 | | 282 | 293 | -11 | -3.8% |
| | | | | | | **MAPE** | 8.7% |
| | | | | | | **MALPE** | 1.0% |

Source: Thorn (2004)

Note, the number of teams (26) was constant between 1980 and 1982

[a] $(P_{0,t} / (P_{1,t} + P_{2,t}) + P_{0,t-1} / (P_{1,t-1} + P_{2,t-1})) \times 0.5$ Years 0

$P_{x,t} / P_{x-1,t-1}$ Years 1 to 9

$P_{10+,t} / P_{9+,t-1}$ Years 10+

[b] $CCR_{0,t} \times (P_{1,t+1} + P_{2,t+1})$ Years 0

$CCR_{x,t} \times P_{x,t}$ Years 1 to 9

$CCR_{9+,t} \times P_{9+,t}$ Years 10+