S.A. Journal for Research in Sport, Physical Education and Recreation, 2002, 24(2): 37-44.
S.A. Tydskrif vir Navorsing in Sport, Liggaamlike Opvoedkunde en Ontspanning, 2002, 24(2): 37-44.
ISSN: 0379-9069

THE COMBINED BOWLING RATE AS A MEASURE OF BOWLING PERFORMANCE IN CRICKET

Hermanus H. LEMMER
Department of Mathematics and Statistics, Rand Afrikaans University, Johannesburg, Republic of South Africa

ABSTRACT

A single measure that can be used to assess the performance of bowlers in cricket is defined. This study shows how it can be used to rank bowlers. The performance of bowlers is generally measured by using three different criteria, i.e. the average number of runs conceded per wicket taken (A), the economy rate (E), which is the average number of runs conceded per over bowled, and the strike rate (S), which is the average number of balls bowled per wicket taken. Each of these is important in its own right. The average (A) is normally used to rate bowlers. This classification is however not very accurate as it does for example not take into account how many overs have been bowled. Two bowlers might have the same average but one may be more economical than the other. The purpose of this paper is to introduce a single measure that takes the full performance of a bowler into account. The combined bowling rate CBR = 3/[1/A + 1/E + 1/S] *is defined for this purpose and is used to rate bowlers. A classification scheme . with ten classes according to which bowlers can be classified is given. The best bowlers are those who fall in class one.*

Key words: Bowlers; Bowling performance; Classification scheme; Combined bowling rate (CBR); Cricket.

INTRODUCTION

Bowlers in cricket are judged according to three criteria, namely the average (A), the economy rate (E) and the strike rate (S) (ODI Bowling, 2001). Each one of these is important in its own right, but it may be useful to define a joint measure of bowling performance in order to assess bowlers' overall performance. Such a measure is proposed and discussed in this article.

Let O = number of overs bowled by a bowler, B = number of balls bowled, R = number of runs scored off his bowling and W = number of wickets taken by the bowler. Then the average A = R/W is a criterion often used to rate bowlers (Waite, 1961:156-157). Also important, though, are the economy rate E = R/O and the strike rate S = B/W. These can be calculated for a bowler's career and per innings, except when the bowler takes no wickets, in which case A and S are not defined. Each of these criteria is a *rate* and each should ideally be as small as possible.

Before attempting to define a combined rate, the magnitudes of each were examined. For this purpose, a data set consisting of the test career bowling figures of all the current test players who have bowled at least one hundred overs was taken from the Internet (CricInfo, 2001). The values of O, R and W were given in the data set, but B had to be calculated using B = 6 x O. Note that no-balls are not taken into account. According to Rob Eastaway (2002), who is actively involved in the PriceWaterhouseCoopers ratings, the strike rate is "the number of legitimate balls per wicket". This is logical because a bowler cannot claim a wicket from a no-ball. The means and standard deviations of A, E and S are given in Table 1.

TABLE 1. **DESCRIPTIVE STATISTICS OF CURRENT TEST BOWLERS' DATA SETS**

	A (Average)	**E (Economy rate)**	**S (Strike rate)**
Mean	35.46582	2.85129	75.27565
Standard deviation	11.05808	0.39227	22.89179

The distributions of A, E and S are all slightly positively skewed - like gamma distributions, but with markedly different means. It is clear that a joint criterion defined as the arithmetic mean of A, E and S will not be a good measure, because it will be dominated by the value of S while the E value will have almost no influence. Some form of standardization is therefore required.

The first form of standardization is as follows: Denote the mean of A by AM and its standard deviation by AS, and similarly for E (EM and ES) and S (SM and SS). The standardized values are TA = (A – AM)/AS, TE = (E – EM)/ES and TS = (S – SM)/SS. The arithmetic mean between these is therefore

$$
\begin{aligned}
AM &= (TA + TE + TS)/3 \\
&= [(A - 35.5)/11.06 + (E - 2.9)/0.3923 + (S - 75.3)/22.89]/3 \\
&= 0.8945(0.0337A + 0.9500E + 0.0163S - 5.1292) \\
&= 0.8945(K - 5.1292)
\end{aligned}
$$

where $K = 0.0337A + 0.9500E + 0.0163S$

Besides constants, AM is a weighted mean between A, E and S with E having the dominant weight as reflected in K. K is actually the statistic that can be used to compare bowlers, because its ratings will always agree with those of AM. This follows from the mathematical relationship between K and AM.

A second method of standardization, which is useful in the case of positive distributions, is to divide each value by its mean, which serves as a scale parameter (Casella & Berger, 1990:118). Such standardized values then have a mean of one.
Let:

$$AG = A/AM, \quad EG = E/EM \text{ and } SG = S/SM$$

Now let:

$$
\begin{aligned}
RG &= (AG + EG + SG)/3 \\
&= 0.1307(0.0719A + 0.8942E + 0339S) \\
&= 0.1307L \text{ with}
\end{aligned}
$$

$$L = 0.0719A + 0.8942E + 0.0339S \text{ the essential statistic.}$$

If we compare L and K, we would prefer L, because the weight of E is smaller in the latter, giving a better balance between A, E and S than in the case of K.

It is not surprising that a logical combined rate would be in the form of a weighted mean between A, E and S. Note that the difference between K and L lies only in the weights. These weights have been arrived at by rational means. It is of no use going into a debate about the choice of weights because there is an unlimited number of possibilities and it is unlikely that any two people will agree on exactly the same choice.

After having looked at two possible statistical approaches, the question of appropriate weights will be answered by using a more basic approach. The question of whether the formula proposed for tests will also be suitable for limited overs matches will have to be addressed. The possible *weighting* of the wickets taken by a bowler will also be discussed.

METHODS

According to Kenney and Keeping (1954:57), the *harmonic mean* is used to find the average of ratios x/y, such as rates, if the unit of the numerator can be considered as fixed and the denominator as variable. Croxton *et al*. (1968:184) state that "the harmonic mean may be useful when data are customarily or conveniently given in terms of problems solved per minute, miles covered per hour, units purchased per dollar, and so forth". See also Iman (1983:575) for more information on the harmonic mean. The rates to be combined in these references are all of the same type. On the other hand, this study wants to combine different types of rates, namely R/W, R/O and B/W. By using the *same principles* as those used in the references for the combination of rates, start by looking at A = R/W and E = R/O. The numerator is the same. One could ask: "If the bowler has conceded R runs, how many overs have been used and how many wickets have been taken?" The harmonic mean is now defined:

$$H = 1/[(1/A + 1/E)/2] = 2/[W/R + O/R] = 2R/[W + O]$$

To judge whether this measure makes sense it is rewritten as

$$H = (R + R)/[W + O] = (W.A + O.E)/[W + O]$$

from which one can see that it is simply a weighted mean between A and E. Clearly, E will generally have a much higher weight than A, which makes much more sense than using an ordinary (unweighted) mean. If one looks at the product of each weight and measure, namely W.A and O.E, one notices that W.A = R and O.E = R, so these weights have the effect of letting A and E contribute equally to H. The harmonic mean is thus a sensible combination between A and E.

The question that remains is: What about the strike rate S? A measure that takes not only A and E into consideration, but also S is needed. The same procedure as before is followed by defining the combined bowling rate as the harmonic mean between A, E and S:

$$CBR = 3/[1/A + 1/E + 1/S]$$

This is a weighted mean between A, E and S because

$$CBR = 3/[W/R + O/R + W/B] = 3R/[W + O + W.R/B]$$

$$= (R + R + R)/[W + O + T] \quad \text{where } T = W.R/B$$

$$= (W.A + O.E + T.S)/[W + O + T]$$

Again the new term is such that weight x rate = T.S = R just as in the case of W.A and O.E.

In order to get a closer look at typical weights, suppose one has a bowler whose performance is the same as the average of the data set. If he has bowled say O = 25 overs, then one has approximately R = 71.28 and W = 1.98 so that T = 0.9409 and therefore the "typical" bowler has:

$$CBR = (1.98A + 25S + 0.94S)/[1.98 + 25 + 0.9409]$$

$$= 0.0709A + 0.8954E + 0.0337S$$

The weights agree very well with those in L, which serves as a further justification for using CBR as an overall bowling criterion. It was mentioned that the rates A, E and S should ideally be as small as possible, so obviously CBR should also be as small as possible.

CBR is self-weighting in the sense that the weights of A, E and S depend on the bowler's statistics (the weights are not fixed numbers determined by a large group of bowlers, as in L). CBR is preferred to L, because firstly, its construction is logical and based on the principles used in the combination of rates. Secondly, if W = 0, L will be undefined but CBR still works well. Thirdly, L depends on a data set and its weights will vary every time the data is updated. The weights in L will furthermore be different for test matches, one-day internationals and different kinds of domestic matches. On the other hand, the formula for CBR remains exactly the same. For calculation purposes, it is easiest to use the formula in the form of

$$CBR = 3R/(W + O + W.R/B)$$

The group consisting of all current One-day International (ODI) bowlers taken on the same date as the test bowlers included 123 who have bowled at least one hundred overs each. For the data set consisting of the ODI career bowling figures of these bowlers, exactly the same procedure as above has been followed, and the following results were obtained:

$$K = 0.0469A + 0.9110E + 0.0421S \text{ and}$$

$$L = 0.1085A + 0.8076E + 0.0839S$$

For a "typical" bowler, $CBR = 0.1085A + 0.8078E + 0.0837S$

Note that the weights differ from those in test matches. Note also the agreement between L and CBR.

A very useful property of CBR is that it can be calculated per ball bowled, per match or for any part of a bowler's career, and it makes sense in every situation.

RESULTS AND DISCUSSION

The numerical value of CBR has no specific physical meaning so it is difficult to judge a bowler's performance by just looking at the value. In order to assess a bowler's performance based on his CBR value, the data set of this study was used to construct ten classes according to which bowlers are classified. Those ten per cent with the smallest (best) CBR values were classified into class one, the next ten per cent into class two, etc. In order to determine the class boundaries once and for all, one must have an idea of the distribution of the CBR. For both data sets, the gamma distribution (Casella & Berger, 1990:100) provided the best fit, but with different parameters for tests and ODIs because the conditions and rules of ODIs and tests are different. The most important statistics are given in Table 2.

TABLE 2. **DESCRIPTIVE STATISTICS OF SETS OF CBR VALUES FOR TESTS AND ODIS WITH THE VALUES OF THE PARAMETERS OF THE GAMMA (a,b) DISTRIBUTIONS THAT GAVE THE BEST FIT**

	Mean	**Std deviation**	**Scale parameter (b)**	**Shape parameter (a)**
Tests	7.5766	0.9788	0.1246	60.8073
ODIs	11.1871	1.1423	0.1158	96.5749

Although the gamma distributions fitted the data statistically well, graphic examinations of the fits revealed that in some regions, the fits were not good. It will therefore not be satisfactory to use the deciles of the fitted gamma distributions as class boundaries. The bootstrap technique (Efron, 1990:79) is rather used to estimate the deciles for the data sets, and the following class boundaries are recommended (see Table 3):

TABLE 3. **A CLASSIFICATION SCHEME OF TEN CLASSES FOR CBR VALUES OF TESTS AND ODIS**

Class number	**Interval for tests**	**Interval for ODIs**
1	0.00 - 6.36	0.00 - 9.58
2	6.36+ - 6.73	9.58+ - 10.15
3	6.73+ - 7.10	10.15+ - 10.52
4	7.10+ - 7.40	10.52+ - 10.84
5	7.40+ - 7.61	10.84+ - 11.13
6	7.61+ - 7.85	11.13+ - 11.46
7	7.85+ - 8.13	11.46+ - 11.72
8	8.13+ - 8.43	11.72+ - 12.12
9	8.43+ - 8.86	12.12+ - 12.76
10	8.86+ -	12.76+ -

It is recommended that the classes for tests be used in the case of ordinary (unlimited overs) matches and those for ODIs for limited overs matches.

It may be difficult to judge how good a bowler is if he has CBR = 6.15 for his test career, but if one notes that he falls into class one, one knows that he is very good. After having used these values for some time, one will get a feeling for their numerical significance.

In this study, all the bowlers have bowled at least one hundred overs in the specific kind of match. The classes in Table 3 are applicable to bowlers who have bowled at least one hundred overs. To study the performance of a bowler, his up-to-date CBR value against each completed innings was plotted. The CBR value of the first innings is plotted against innings one. After the second innings, the CBR for both innings is calculated and plotted against innings two, and so forth. Each CBR is thus a value based on cumulative totals of overs, runs and wickets. In the first few innings, the CBR fluctuates a lot and for most bowlers it starts off with small values. After about ten innings (in ODIs this can be up to one hundred overs bowled), it stabilizes and then fluctuates around a kind of limiting value that characterizes the bowler. The classification scheme of this study generally favours a new bowler because his CBR will normally start off with a low value, but fairly soon he will find his place in the classification scheme. A bowler's performance can be monitored by looking at his most recent CBR values, or at the class into which he falls.

Instead of using the CBR on a cumulative basis, the CBR for each innings could be calculated and the average, CM, of all these CBRs could then be calculated up to date. It is interesting to note that the career plot of CM has very much the same pattern as that of CBR with values slightly larger than those of CBR. Though they differ numerically, they tell the same story. The calculation of CBR is, however, much simpler than that of CM because it requires only overall bowling figures and not figures per innings. So in this study, the CBR is preferred.

In the calculation of the CBR, it is possible to weight the wickets of top order batsmen higher than those of lower order batsmen, as is done in the PriceWaterhouseCoopers (2002) ratings. In the present study, the mean number of wickets taken per bowler is large, i.e. 78. It is thus reasonable to assume that most bowlers have taken wickets of top and lower order batsmen and that the results of this study would not have differed much if the more complicated weighting procedure had been followed.

APPLICATION

In order to assess the likely performances of teams who are about to enter into a test or ODI series, it can be useful to compare the bowlers. For illustration purposes, the data of the SA, Australian and New Zealand ODI bowlers as at 31 January 2002, i.e. towards the end of their triangular series, was used. In Table 4, the names in order according to CBR are given and their class numbers are mentioned.

TABLE 4. **RANKING OF BOWLERS ACCORDING TO CBR VALUES IN ODIS**

Rank	Name	Age	Overs	Runs	Wickets	CBR	Class
1	Shaun Pollock	28	1297	4855	214	8.857	1
2	Glenn McGrath	31	1311.4	5207	225	9.269	1
3	Allan Donald	35	1217.3	4966	232	9.270	1
4	Shane Warne	32	1604.3	6812	268	9.911	2
5	Makhaya Ntini	24	228.3	1198	43	9.953	2
6	Damien Fleming	31	769.5	3402	134	10.183	3
7	Jason Gillespie	26	261.1	1130	38	10.383	3
8	Steve Elworthy	36	275.4	1210	42	10.426	3
9	Daryl Tuffey	23	133	623	25	10.529	4
10	Brett Lee	25	290.4	1393	57	10.634	4
11	Chris Harris	32	1485.3	6381	180	10.669	4
12	Nicky Boje	28	404.5	1780	51	10.834	4
13	Jacques Kallis	26	828.5	3830	128	10.890	5
14	Lance Klusener	30	914	4312	151	10.928	5
15	Ian Harvey	29	306.4	1394	42	10.998	5
16	Andy Bichel	31	185.5	869	28	11.077	5
17	Daniel Vettori	23	612.4	2737	72	11.126	5
18	Chris Cairns	31	990.2	4679	141	11.300	6
19	Nathan Astle	30	688.1	3179	86	11.349	6
20	Andy Symonds	26	263.2	1299	43	11.409	6
21	Mark Waugh	36	614.3	2938	85	11.491	7
22	Dion Nash	30	558.2	2579	61	11.614	7
23	James Franklin	21	147	720	19	11.900	8
24	Craig McMillan	25	207.3	1091	36	11.907	8

CONCLUSION

It was previously mentioned that each of the three criteria A, E and S is important in its own right. The tendency is, however, to compare or rate bowlers by looking at the average A only. The problem of just using A is that it takes R and W into account, but not O. One can thus have a very economical bowler with exactly the same average as a very uneconomical bowler and rate them as equally good. This study strongly recommends that this is not good enough and that a measure like CBR could fill an important gap in the assessment and rating of bowlers.

REFERENCES

CASELLA, G.C. & BERGER, R.L. (1990). *Statistical Inference.* Pacific Grove: Wadsworth & Brooks.

CRICINFO (2001). Test career bowling averages. [http://www-rsa.cricket.org/link_to_database/STATS/]. Retrieved 28 July 2001.

CROXTON, F.E.; COWDEN, D.J. & KLEIN, S. (1968). *Applied General Statistics* (3rd ed.). London: Pitman & Sons.

EASTAWAY, R. (2002). [rob@eastaway.demon.co.uk]. Strike rate. Private e-mail message to Hermanus Lemmer [hhl@na.rau.ac.za]. 23 July 2002.

EFRON, B. (1990). More efficient bootstrap computations. *Journal of the American Statistical Association,* 85(409): 79-89.

IMAN, R.L. (1983). Harmonic Mean. In S. Kotz & N.L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, 3 (575-576). New York, NJ: John Wiley & Sons.

KENNEY, J.F. & KEEPING, E.S. (1954). *Mathematical Statistics* (3rd ed.). New York, NJ: D. van Nostrand.

ODI BOWLING (2001). Career summary. [http://statserver.cricket.org/perl/sdb/sdb_find.pl/]. Retrieved 20 July 2001.

PRICEWATERHOUSECOOPERS (2002). Cricket ratings. [http://cricketratings.pwcglobal.com/cricket/cricket.htm]. 15 January 2002.

WAITE, J. (1961). *Perchance to bowl.* London: Nicholas Kaye.

Prof. Hermanus H. Lemmer, 122 Fourth avenue, Fairlands 2195, Republic of South Africa. Tel. (w) 27 (0)11 678-4413, Tel. (h) & Fax. 27 (0)11 678-3031, Cell. 073 1661934, E-mail: hhl@na.rau.ac.za

(Subject editor: Prof. H. de Ridder)