AN OBJECTIVE METHOD FOR SETTING ASA NATIONAL AGE GROUP QUALIFYING TIMES

A report prepared for the ASA Technical Swimming Committee
by
Dr. G R Thwaites Ma, M.Eng, Ph.D.

Introduction

This paper describes work carried out to develop a fully objective method of setting National Age Group and Youth Championship qualifying times. The work was carried out with the support and cooperation of the other members of the Qualifying Times Setting Committee (in particular Barry Saunders, Chris Bostock and Graham Sykes).

Summary

A method has been developed to set National Qualifying Times for Age Group and Youth Championships, both long and short course. The method is fully objective and depends only on the FINA 1000 point standard and the GB all-time best ranking database. No adjustments are necessary either for age or event. Extrapolation may be necessary in a small number of event/age combinations where there are few entries in the database. For 2005, the criterion for the Youth Championships differed slightly from the Age Groups because entries were limited by the available time, although both sets of times were calculated using the same basic method. It has been recommended that single age group qualifying times should be used for the Youth Championships.

Objectives

The objectives of this work were as follows:

- To develop a completely objective method of setting National Age Group and Youth Championship qualifying times
- To maintain a connection between these qualifying times and world standards.
- That a single method should apply to both Age Group and Youth Championships and provide both long and short course qualifying times.
- That the new method should give times roughly in line with the 2004 qualifying times.
- A subsidiary objective was that the chosen method could be modified in a systematic manner to give National Senior, District and Graded meet qualifying times.

Methods

The methods described below refer to the ‘age progression curves’. What follows will probably be more clear if the previous method for setting age group qualifying times is explained:

1. For each age/event/gender combination the top ten in the ranking list are converted to GB points using the formula given in (1) below.
2. The average value of these ten points values is taken.

3. For each age/gender combination, the event having the maximum points average is chosen. This value is known as the ‘pivot point’. A graph of the pivot points against age (for each gender) is known as the INT progression curve.

4. An ‘Age Correction Factor’ (ACF) is calculated for each event/age/gender combination by dividing the average of the top ten points values for that combination by the pivot point for the equivalent age/gender combination. These values will all lie below 1 (as the pivot point was the maximum).
5. An ‘AAA’ or ‘100%’ progression curve is calculated by taking a fixed percentage of the difference between the INT progression curve and a notional ‘B’ curve. The B curve has been generated by G.Sykes to represent the average club level swimmer and is based on observations over many years.
6. The points value for a given age from the 100% curve is multiplied by each ACF to give a qualifying points value (p_q) for each event. This is then converted to a time (t_q) using the formula given in (3) below. The values of t_q are used as the basis for the National Age Group qualifying times. Some adjustments have usually been made for individual events where it was thought necessary.

The National Youth Championship qualifying times have recently been calculated using a ‘countdown’ system, whereby, for example, the qualifying time for each event/age/gender combination has been set as the 75 ranking in the ATB list. This has also needed intervention to modify times, especially in the less popular events where there are relatively few entries in the ATB lists.

The above methods are not completely objective partly because some values have historically been adjusted and because the B curve values are subjective.

The source data used for the work reported here was the ASA 2003 ‘All time’ best performance tables (here referred to ATBL for long course and ATBS for short course). These in fact contain the best performance data from 1995 to 2003 – 9 years.

To follow the previous system closely and to maintain the links with world standards, calculations were based on GB points (GBP) for each event. GB points are calculated from the swim time t as follows:

$$GBP = 10^{(4-t/t_k)} \qquad (1)$$

Where t_k is the FINA 1000 point reference time for the event in question (at the time of calculation based on the average of the world all time top 8). A time of t_k for any event gives 1000 GBP and twice this time gives 100 GBP. This method was developed by G.Sykes and has been used for a number of years. It represents a transformation of the data which normalises each event based on the top world standard for that event, so gives some degree of comparison between events.

Initial work was carried out to look at the distribution of actual times for a number of events. The object of this work was to see whether a function could be fitted to the distribution for each event so that any point on the distribution could be extracted from a formula. The methods and results of this work are given in Appendix 1. Investigating the nature of this data was valuable in terms of deciding the method of using the ATB tables in the calculation methods studied. To summarise, it was found that a truncated normal distribution gave an excellent fit to the top half of the distribution but did not generally fit the top tail well, especially the top 10 – 15 points.

It was found not to be necessary to use the fitted distribution as the data was so good in the area of interest that it was sufficient to use a small range average of the actual GBP data for each event. The reference value chosen was the average of ranks 15 to 25 (this covers 11 data points but is designed to represent the average at the 20th rank). Generally this value agreed with the best fitted normal distribution to within a few one hundredths of a second. This value (expressed in GBP) is referred to as p_{20} below).

The reason for choosing the 15-25 range (rather than places 1-10 as previously) was that, as mentioned above, the very top of the distribution was not always representative of lower values, often being dominated by a small number of exceptional swimmers. This is discussed further in Appendix 1.

Three methods of generating qualifying times were studied:

1. 'Ratio method': This used a fixed percentage of the GBP 15-25 average reference value (p_{20}) for each age/gender/event combination to set a GBP qualifying value (p_q) which can then be converted back into an actual time (t_q):

$$p_q = k \times p_{20} \quad (2)$$

$$t_q = t_k \times \{ 4 - \log_{10} (p_q)\} \quad (3)$$

The value of k to be used was selected as an average over all events so that the age group progression curve matched that used to set the 2004 age group qualifying times as closely as possible. Separate values of k were used for long and short course but a single value was used for all event/age/gender combinations. This means that the qualifying times for all events are set on exactly the same basis.

As an example, if the value for k for the long course events was chosen as 0.9, the qualifying time for each event would be based on a GBP value of 0.9 x the average of ranks 15-25 in that event.

2. 'Count method': This used a simple counting method down the ATB table for each event. This is similar to that used in 2004 for the Youth Championships (although the times were adjusted individually where it was thought necessary). Once again, a single criterion was used for all event/age/gender combinations but separate values for long and short course. The value was again chosen to give the best match to the age progression curves used for the 2004 age group qualifying times.

As an example, if rank 150 was chosen for long course, the qualifying time for each event would be set at rank 150 in the ATBL table for that event. This method needs no calculation of GBP.

3. 'Rank percent method': A percentile of the distribution for each event is chosen. A range of +/- 5 percentage points around the given percentile is chosen to smooth the data in the same way as for the Ratio Method. This average is used as the qualifying points value for each event/age/gender combination (p_q). Once again a single criterion is used for all events but separately for long and short course.

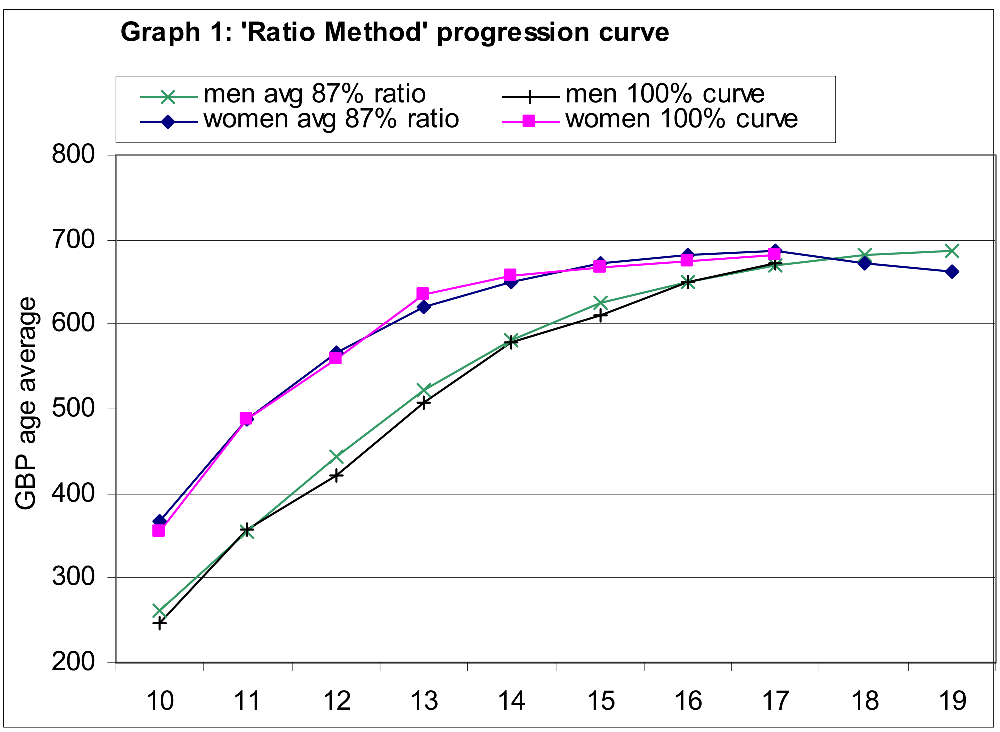
The percentile to be used is chosen to match the 2004 age progression curves as closely as possible.

As an example, if the top 15th percentile is chosen for the long course events and a given event has 1000 entries in the data base, the GBP for ranks 100 to 200 (top 10th to 20th percentile) in the ATBL are averaged to give p_q and the qualifying time (t_q) is calculated from p_q as for the Ratio Method.

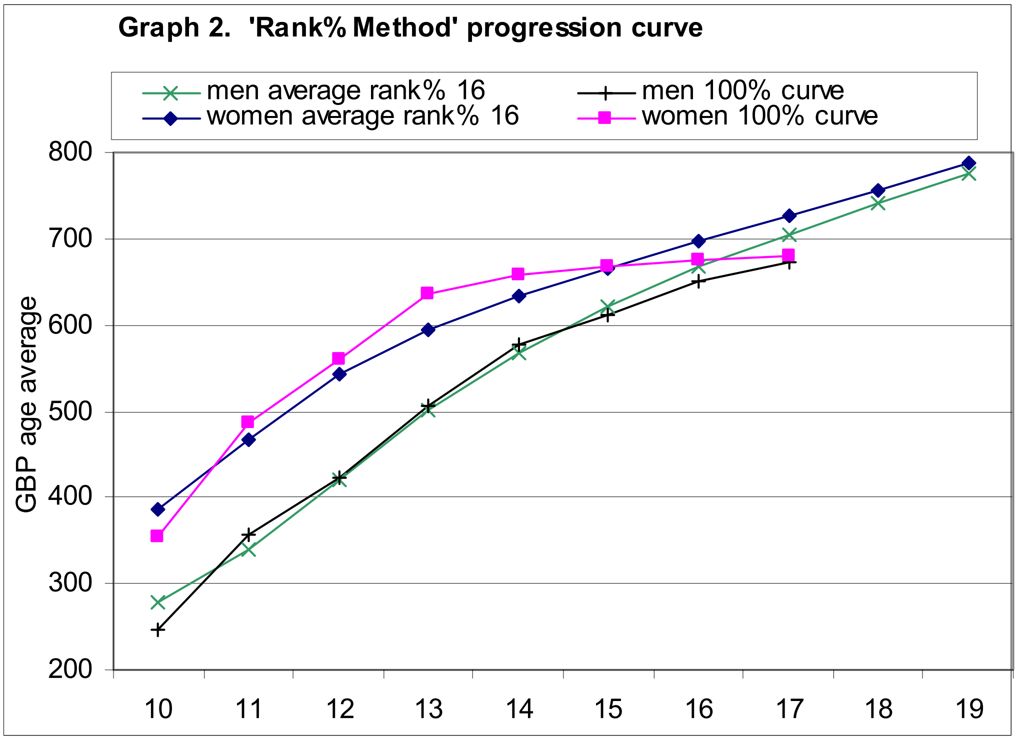
Results

Graphs 1, 2 and 3 compare the qualifying points values, averaged over all events for each age/gender combination for each method with the 2004 equivalent 100% age progression curves for long course times (this was the value used to set the 2004 National Age Group qualifying times).

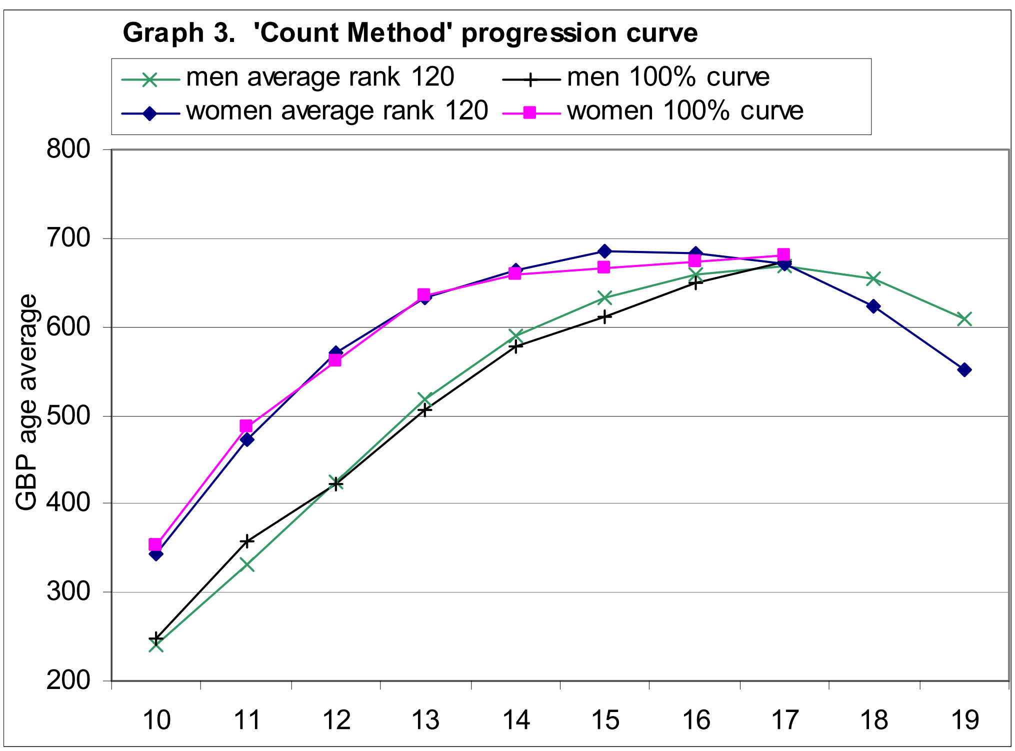
A value of k of 0.87 was chosen for the ratio method to give the best overall fit to the
existing method.



Rank percent 16 (average of rank percents 11 to 21) was chosen for the rank percent method to give the best overall fit to the existing method.



Rank 120 was chosen for the count method to give the best overall fit to the existing method.



The actual qualifying times predicted for each method (based on the 2003 ATB lists) are compared with the actual 2004 qualifying times in Table 1.

Table 1. Comparison of qualifying times set by new calculation methods with 2004 NQT (Long Course)

RED	indicates a time which is more than 2% slower than 2004 time
BLUE	indicates a time which is more than 2% faster than 2004 time
	indicates there is is too little data in the ATB list to set a meaningful time

	11							12							13						
	NQ04	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**	
		Time	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**
M50FS		31.83		31.78		32.60			30.10		30.04		30.87			28.40		28.66		28.91	
M100FS		68.35		68.33		70.32		64.65	63.82	-1.3	63.89	-1.2	66.45	2.8	61.68	60.35	-2.2	60.95	-1.2	62.59	1.5
M200FS	151.12	149.75	-0.9	148.47	-1.8	150.75	-0.2	140.74	139.07	-1.2	139.90	-0.6	141.85	0.8	134.44	131.51	-2.2	132.85	-1.2	134.45	0.0
M400FS	312.19	312.52	0.1	307.00	-1.7	309.48	-0.9	295.65	291.17	-1.5	291.55	-1.4	293.76	-0.6	282.75	277.81	-1.7	279.07	-1.3	280.24	-0.9
M1500FS						1256.8		1176.9							1121.8						
M50BR		42.35		41.55		43.06			39.35		38.41		39.88			36.59		36.14		36.69	
M100BR		90.66		88.72		92.27		82.17	83.26	1.3	81.79	-0.5	85.93	4.6	77.66	77.95	0.4	77.41	-0.3	80.56	3.7
M200BR	194.85	197.50	1.4	192.15	-1.4	195.03	0.1	178.21	181.27	1.7	178.73	0.3	181.51	1.8	167.71	168.55	0.5	168.46	0.4	170.49	1.7
M50FL		35.90		35.27		36.52			33.58		32.95		34.07			31.51		31.16		31.76	
M100FL		79.45		77.13		79.71		71.00	73.12	3.0	71.71	1.0	74.89	5.5	66.76	68.18	2.1	67.31	0.8	70.13	5.1
M200FL	174.37	188.34	8.0	173.89	-0.3	170.72	-2.1	157.97	166.64	5.5	159.52	1.0	159.44	0.9	149.56	152.70	2.1	149.81	0.2	150.55	0.7
M50BA		37.58		36.99		38.26			35.42		35.09		36.02			33.50		33.24		33.64	
M100BA		79.28		77.72		80.96		72.95	73.20	0.3	72.87	-0.1	76.38	4.7	69.77	69.35	-0.6	69.24	-0.8	72.27	3.6
M200BA	170.90	170.42	-0.3	166.97	-2.3	169.18	-1.0	158.29	159.06	0.5	157.32	-0.6	159.66	0.9	149.71	149.39	-0.2	149.29	-0.3	151.47	1.2
M200IM	171.97	169.64	-1.4	168.05	-2.3	172.22	0.1	159.77	158.16	-1.0	157.80	-1.2	162.59	1.8	152.41	149.87	-1.7	150.09	-1.5	153.49	0.7
M400IM	370.67			372.53	0.5	343.49	-7.3	339.75	352.00	3.6	338.48	-0.4	327.61	-3.6	323.12	320.99	-0.7	320.21	-0.9	317.53	-1.7
av change			1.2		-1.3		-1.6			1.0		-0.3		1.8			-0.4		-0.5		1.4

	14							16							18						
	NQ04	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**	
		Time	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**
M50FS		27.15		27.56		27.58		25.72	25.80	0.3	26.19	1.8	26.14	1.6	25.32	25.49	0.7	25.62	1.2	25.33	0.0
M100FS	59.43	58.00	-2.4	59.10	-0.6	60.09	1.1	55.50	55.94	0.8	56.80	2.3	57.01	2.7	54.83	55.61	1.4	55.95	2.1	55.08	0.5
M200FS	129.13	126.72	-1.9	128.71	-0.3	129.63	0.4	120.74	121.80	0.9	123.93	2.6	122.82	1.7	119.36	122.06	2.3	122.01	2.2	118.48	-0.7
M400FS	271.84	267.56	-1.6	271.81	0.0	270.58	-0.5	257.72	259.27	0.6	260.21	1.0	257.98	0.1	253.92	263.40	3.7	259.61	2.2	248.86	-2.0
M1500FS	1078.6			1070.8	-0.7	1053.1	-2.4	1010.1	1050.5	4.0	1038.4	2.8	1001.6	-0.8	1003.9			1040.3		969.4	
M50BR		34.87		34.89		34.83			33.16		33.43		32.76			32.58		32.26		31.35	
M100BR	74.47	74.19	-0.4	74.67	0.3	76.93	3.3	70.88	71.41	0.8	72.06	1.7	72.15	1.8	69.85	71.37	2.2	70.73	1.3	68.64	-1.7
M200BR	161.33	161.49	0.1	161.71	0.2	163.35	1.3	155.42	156.71	0.8	156.33	0.6	155.11	-0.2	153.57	159.94	4.2	153.30	-0.2	148.49	-3.3
M50FL		29.88		30.00		30.04			28.24		28.51		28.25			27.73		27.83		27.22	
M100FL	64.53	64.76	0.4	64.83	0.5	66.72	3.4	60.92	61.37	0.7	61.68	1.2	62.02	1.8	59.97	61.00	1.7	60.96	1.6	59.54	-0.7
M200FL	142.53	145.15	1.8	143.19	0.5	143.42	0.6	136.59	138.51	1.4	136.70	0.1	134.22	-1.7	134.16	142.32	6.1	134.46	0.2	128.35	-4.3
M50BA		31.94		31.85		32.01			30.46		30.48		30.21			30.28		29.50		28.97	
M100BA	66.21	66.39	0.3	66.55	0.5	69.09	4.3	63.50	63.87	0.6	64.13	1.0	64.67	1.8	62.77	64.45	2.7	63.04	0.4	61.72	-1.7
M200BA	143.38	142.76	-0.4	143.44	0.0	144.72	0.9	136.93	138.18	0.9	138.16	0.9	136.99	0.0	135.80	141.02	3.8	136.43	0.5	131.09	-3.5
M200IM	145.96	143.79	-1.5	145.64	-0.2	147.69	1.2	137.94	138.66	0.5	140.12	1.6	139.45	1.1	136.61	138.96	1.7	137.41	0.6	134.05	-1.9
M400IM	309.78	306.80	-1.0	309.53	-0.1	305.17	-1.5	294.35	296.84	0.8	297.12	0.9	290.27	-1.4	291.91	303.26	3.9	293.15	0.4	278.83	-4.5
av change			-0.6		0.0		1.0			1.0		1.4		0.7			2.9		1.0		-2.0

	11							12							13						
	NQ04	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**	
		Time	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**
W50FS		32.02		32.04		32.91			30.59		30.94		31.50			29.76		30.12		30.50	
W100FS	69.25	68.95	-0.4	68.68	-0.8	71.11	2.7	66.15	64.94	-1.8	65.71	-0.7	67.95	2.7	64.16	63.15	-1.6	64.37	0.3	65.73	2.4
W200FS	150.49	149.87	-0.4	149.76	-0.5	150.75	0.2	142.97	141.12	-1.3	142.13	-0.6	144.12	0.8	137.92	136.46	-1.1	138.15	0.2	139.67	1.3
W400FS	311.59	312.56	0.3	312.85	0.4	309.79	-0.6	297.71	294.43	-1.1	296.67	-0.3	297.83	0.0	287.77	285.55	-0.8	289.34	0.5	289.33	0.5
W800FS	652.53			656.89	0.7	641.18	-1.7	614.6	618.96	0.7	615.39	0.1	601.59	-2.1	588.1	589.64	0.3	594.80	1.1	586.92	-0.2
W50BR		41.65		41.10		42.91			39.47		39.11		40.44			38.09		37.97		38.78	
W100BR	88.03	89.05	1.2	87.74	-0.3	91.81	4.3	83.67	83.10	-0.7	83.65	0.0	87.52	4.6	79.74	80.92	1.5	80.54	1.0	84.48	5.9
W200BR	189.00	194.17	2.7	190.55	0.8	193.41	2.3	180.74	180.25	-0.3	181.08	0.2	182.43	0.9	171.76	173.80	1.2	173.45	1.0	177.34	3.2
W50FL		35.93		35.28		36.71			33.87		33.74		34.58			32.65		32.41		33.30	
W100FL	76.73	79.31	3.4	77.15	0.5	81.02	5.6	72.40	73.21	1.1	73.11	1.0	76.38	5.5	70.03	70.47	0.6	70.79	1.1	73.55	5.0
W200FL	172.04	186.30	8.3	173.98	1.1	171.73	-0.2	160.90	165.92	3.1	162.93	1.3	161.27	0.2	153.60	156.14	1.7	155.13	1.0	155.96	1.5
W50BA		37.31		37.02		38.33			35.52		35.61		36.29			34.48		34.87		34.90	
W100BA	78.22	78.78	0.7	78.10	-0.2	81.02	3.6	74.28	73.91	-0.5	74.44	0.2	77.33	4.1	71.97	71.75	-0.3	72.61	0.9	74.94	4.1
W200BA	169.47			168.32	-0.7	167.99	-0.9	158.88	159.43	0.3	159.85	0.6	161.51	1.7	153.90	153.43	-0.3	155.59	1.1	156.88	1.9
W200IM	170.12	168.42	-1.0	168.74	-0.8	172.50	1.4	161.67	158.70	-1.8	161.18	-0.3	164.44	1.7	156.21	154.44	-1.1	157.04	0.5	159.33	2.0
W400IM	363.29			367.46	1.1	346.28	-4.7	342.99	344.40	0.4	341.41	-0.5	332.56	-3.0	327.65	327.03	-0.2	330.42	0.8	326.60	-0.3
av change			1.6		0.1		1.0			-0.2		0.1		1.4			0.0		0.8		2.3

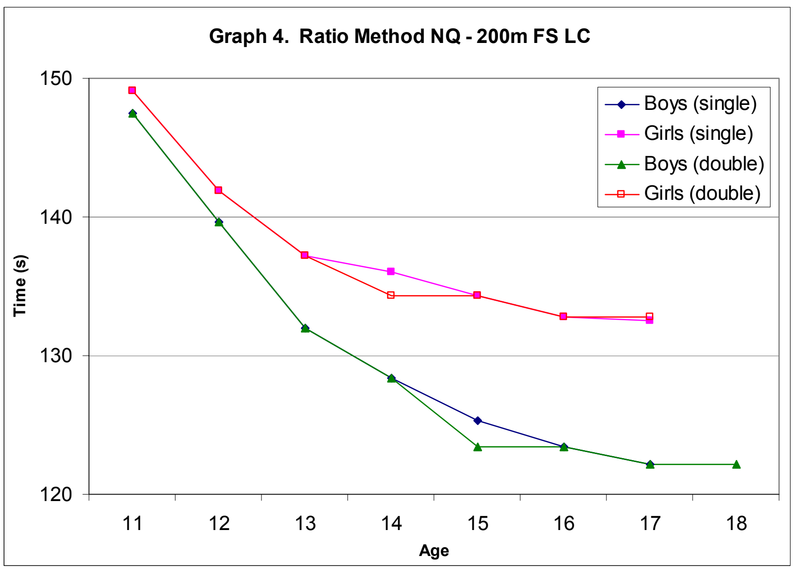
	15							17						
	NQ04	**Count**		**Ratio**		**Rank%**		**NQ04**	**Count**		**Ratio**		**Rank%**	
		Time	**%ch**	**Time**	**%ch**	**Time**	**%ch**		**Time**	**%ch**	**Time**	**%ch**	**Time**	**%ch**
W50FS	28.45	28.61	0.6	29.12	2.3	29.25	2.8	28.37	28.61	0.8	29.00	2.2	28.53	0.6
W100FS	61.03	62.00	1.6	63.27	3.7	64.45	5.6	60.74	61.52	1.3	62.17	2.3	61.28	0.9
W200FS	131.67	132.67	0.8	134.55	2.2	134.75	2.3	131.04	132.74	1.3	132.87	1.4	130.39	-0.5
W400FS	276.27	277.79	0.5	282.84	2.4	279.24	1.1	276.08	280.41	1.6	278.92	1.0	270.92	-1.9
W800FS	569.4	575.42		580.70	2.0	566.00	-0.6	572.3	596.61	4.3	576.74	0.8	548.30	-4.2
W50BR		36.62		36.49		36.83			36.93		36.60		35.76	
W100BR	77.96	78.89	1.2	78.81	1.1	80.75	3.6	78.09	79.59	1.9	78.71	0.8	78.05	-0.1
W200BR	168.91	170.28	0.8	170.02	0.7	170.69	1.1	168.49	173.20	2.8	169.89	0.8	166.30	-1.3
W50FL		31.26		31.36		31.62			31.20		31.20		30.74	
W100FL	67.55	68.31	1.1	68.84	1.9	69.99	3.6	67.29	68.54	1.9	67.84	0.8	67.17	-0.2
W200FL	149.77	151.44	1.1	150.65	0.6	149.15	-0.4	149.19	153.56	2.9	148.89	-0.2	143.81	-3.6
W50BA		33.15		33.29		33.40			33.18		32.96		32.44	
W100BA	69.06	69.61	0.8	70.15	1.6	71.82	4.0	68.76	70.18	2.1	69.67	1.3	69.20	0.6
W200BA	148.09	149.38	0.9	149.47	0.9	150.79	1.8	147.76	151.20	2.3	148.49	0.5	145.36	-1.6
W200IM	149.71	150.32	0.4	152.56	1.9	153.48	2.5	149.29	151.12	1.2	150.63	0.9	148.21	-0.7
W400IM	316.25	319.38	1.0	322.49	2.0	316.18	0.0	316.77	322.70	1.9	318.57	0.6	306.33	-3.3
av change			0.9		1.8		2.1			2.0		1.0		-1.2

There is a considerable amount of data presented in the Graphs and Tables above, so the main conclusions which may be drawn from these are summarised below:

- The age progression curves represent an average over all events for each age, so the quality of fit of the three calculation methods does not necessarily represent the quality of fit for the individual events.
- The progression curve for the ‘Ratio Method’ follows both the actual values and the shapes of the curves for the existing qualifying times very closely and is somewhat smoother.
- The progression curve for the ‘Rank % method’ gives a straighter line than the existing progression curves, giving harder qualifying times at the older age groups.
- The progression curve for the ‘Count Method’ does the opposite, being more downwardly convex and giving easier qualifying times at the older ages, both relative to the existing times and to the younger age groups.
- The fact that the ‘Ratio Method’ gives similar behaviour to the existing age progression curve does not indicate that it is in any way ‘better’. The chosen shape of the progression curve is a subjective decision based on how it is felt that qualifying times should vary with age. An argument for the times for the older age groups being easier would be that, since many swimmers retire at 16-17, the times should be easier to keep them in the sport. Alternatively, it could be argued that harder times would encourage them to train harder and improve standards. However, one of the stated objectives of this work was that the new system should be roughly in line with the existing system which has proved itself over a number of years. On this basis, the ‘Ratio Method’ seems to be the best choice with the added advantage of somewhat smoother progression curves.
- Table 1 shows the simulated individual event qualifying times as would have been calculated by each of the three methods for 2004 in comparison with the actual qualifying times. Changes of more than 2% are highlighted. Although the differences are not quite as consistent as shown by the age progression graphs, the trends can be seen to be the same and both the Count and Rank Percent methods give a lot of values which differ significantly from the existing qualifying times.
- The individual times for the Ratio Method show a small but significant effect of making the qualifying times a little harder at the younger age group end and a little easier at the older end. It was thought by the committee that these were desirable in view of the number of entries in 2003. The Ratio Method was, therefore chosen as the preferred method of setting future qualifying times.
- The values of the ACF (age correction factors) are not shown here. Those used for the existing (2004) qualifying times show rather erratic behaviour. The equivalent ACF values using the Ratio method are smoother, but not perhaps as smooth as might be desired. This is due to two effects: inadequacies in the ranking tables (there not being enough data, even after 9 years to give completely smooth results) and also to real physiological effects such as the onset of puberty. Whilst some further degree of smoothing of the data could correct for the first of these, it would not be entirely desirable because of the second. In any case, the values should become smoother over the years as more

data is added, so mathematical smoothing would just add a further unnecessary complication.

- All the results and discussion above refer to the long course data (based on the ATBL tables). The same analysis was carried out for the short course data and the conclusions were similar. It was found necessary to use a value of k of 0.91 for short course (as opposed to 0.87 for long course). This higher value is both expected and desirable. It rises from the fact that at World level, more events are swum long course while at National Age Group level, swimmers swim short course relatively more frequently. This means the better times in the World ranking are long course times whilst those in National Age Group rankings tend more often to be short course times, thus a harder relative value needs to be used short course than long course for the National Age Groups to maintain the link to World standards.
- When looking at the changes in the proposed qualifying times which would happen as a result of using the new system, it was notes that the smoothness of the actual qualifying time progression curve was affected substantially by the use of double age groups for the youth championships. This is illustrated in Graph 4 which shows the effect for the 200m freestyle events (other vents behave in a similar manner).



This shows a significant jump in the qualifying time graph for boys at 15 and girls at 14 (there is a similar effect at 17 and 16 respectively but it is much smaller). On the 200m freestyle this actually represents a discontinuity of 2 seconds. Looking at the 2003 entries, it can be seen that there are relatively few from 15 year old boys and 14 year old girls. As these are ages when swimmers may become discouraged and quit the sport if they fail to make a qualifying time, the committee recommended that single age group qualifying times be used for 2005 onwards and this has been implemented. (Although this graph shows the proposed times for the new method, the old method gave exactly the same behaviour).

Discussion

This report deals only with the setting of National Age Group qualifying times. The existing software for setting qualifying times at all other levels (District, County and graded meets) can be modified relatively easily to use this method, so no further work is needed immediately to allow the full range of qualifying times to be set. The only disadvantage of the existing software is that it uses the notional B curve to proceed from the National to lower level times. As this method has worked in the past, there is no reason why it should not continue for the present, although it would ultimately be preferable to calculate these completely from the ATB lists and remove the dependence on the B curve which is the only subjective element remaining in the calculations.

In principle senior qualifying times can be set in a similar manner to the age group times, basing them on an appropriate percentage of the senior 15-25 rank average expressed in GBP.

The development of this method from year to year should be considered. As the rankings are based on the all time best, they should change only slowly with the general improvement of standards, i.e. as new swimmers appear in the top 25, the qualifying times will become a little faster. As there are only 9 years worth of data in the ATB as at 2003-4, there will be a slight change due simply to there being extra years of data (without any change in standards). For example, when the 2004 data is entered, the lists will be approximately 10% bigger so one would expect 2-3 new swimmers to appear in each top 25. This effect will be quite small and will reduce year on year. However, it will be preferable to monitor the qualifying times generated by this method for the first few years to check that any improvements in the qualifying times are genuinely due to improving standards. For example the 2003 25th place in the men's 16 years 100m freestyle has moved down to 29th in the 2004 ATB and the 25th time has improved from 54.01 to 53.91.

Conclusions

An objective method has been developed for setting National Age Group qualifying times. This applies to single age groups for all National Age Group and Youth championships.

The algorithm for setting qualifying times using the method recommended here (ratio method) is given below. The procedure needs to be carried out separately for long and short course.

1. Convert the top 25 times in the all time best list for each event/age/gender combination to GB points as follows:

   $$GBP = 10^{(4-t/t_k)}$$

   Where t is the individual time and t_k is the FINA reference time for the event.

2. Average the GB points for ranks 15 to 25 for each event/age/gender combination.

3. Multiply the average by 0.87 for long course or 0.91 for short course. This gives the qualifying points value (p_q).

Convert this back to a time using:

$$t_q = t_k \times \{ 4 - \log_{10} (p_q)\}$$

where t_k is the FINA reference time as before.

4. t_q is the National Age Group qualifying time for that event/ age /gender combination.

The current qualifying time setting program uses the data in a slightly different way. The algorithm is restated below in the way used in the qualifying time setting program.

1. Convert the top 25 times in the all time best list for each event/age/gender combination to GB points as follows:

$$GBP = 10\text{^}(4-t/t_k)$$

   Where t is the individual time and t_k is the FINA reference time for the event.

2. Average the GB points for ranks 15 to 25 for each event/age/gender combination.

3. Multiply the average by 0.87 for long course or 0.91 for short course. This gives the qualifying points value (p_q).

4. For a given age/gender combination, take the maximum value of p_q over all the events. This gives the pivot point.

5. Divide each value of p_q by the pivot point for that age/gender combination. This gives the Age Correction Factor for each event/age/gender.

6. The qualifying points value for each event/age/gender is calculated by multiplying the pivot point by the Age Correction Factor. (This reverses steps 4 and 5).

Convert this back to a time using:

$$t_q = t_k \times \{ 4 - \log_{10} (p_q)\}$$

where t_k is the FINA reference time as before.

7. t_q is the National Age Group qualifying time for that event/ age /gender combination.

This merely adds two extra steps which are then reversed, but doing the calculation this way is more compatible computationally with the existing program.

Further work

As was mentioned earlier, in order to set qualifying times for District, County and graded meets, a notional 'B' curve is used to interpolate the data in the qualifying times setting program. It would be preferable to replace the B curve with a method based entirely on the all time best lists.

The methods used here can easily be extended to cover senior qualifying times at all levels. These, however, depend on policy decisions such as whether the basis should be an absolute link to World levels, based entirely on UK levels or based on a target entry number.

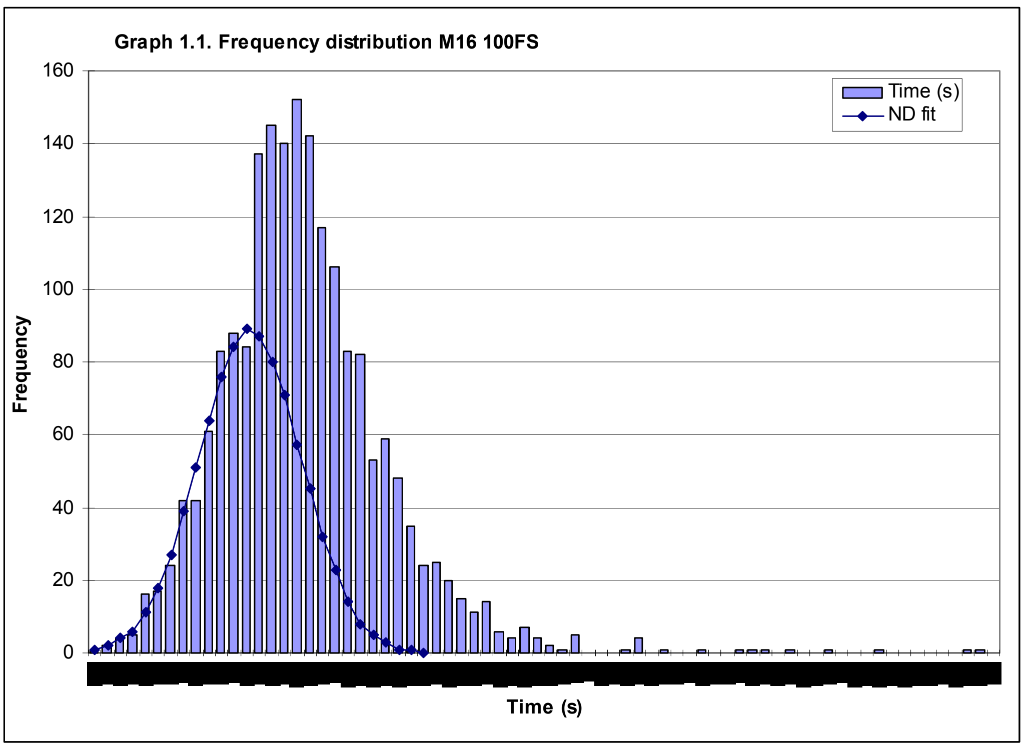
There are other areas where analysis of the all time best lists might be useful such as:

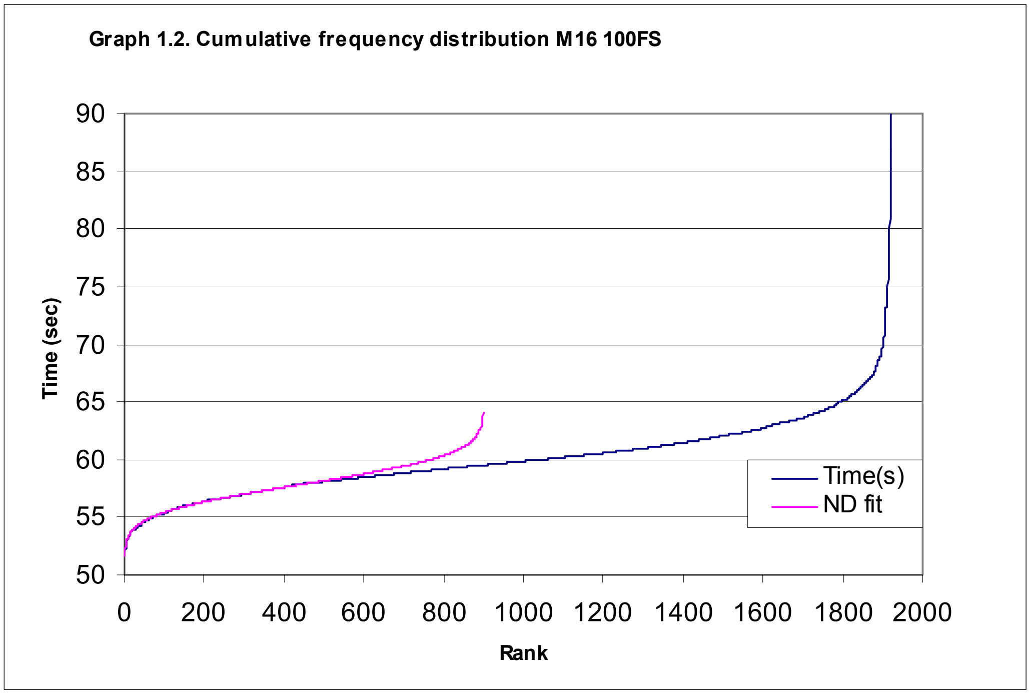
- an alternative method of looking at 25/50m pool equivalent performance.
- developing progression curves to chart, and set targets for, the performance of individual swimmers.
- evaluating the performance of the BAGCATS system in ensuring that age group swimmers develop successfully through to senior level.
- an extension to masters events to facilitate comparison between performances in different age groups and set qualifying times where necessary.

Appendix 1: Distribution of times in ATB lists

As was stated earlier, the work reported in this appendix was not used in the final analysis. However, it is reported here to show the reason for the choice of the average of ranks 15-25 for the reference value and for anyone working in the future on this system. The data presented in this appendix has been updated to use the 2004 ranking lists.

The distribution of the times in the all time best list for a well populated event, the men's 16 years 100 FS (long course), is shown in graphs 1.1 and 1.2. Graph 1.1 is presented as a frequency distribution and graph 1.2 as a cumulative distribution, plotting time against rank.

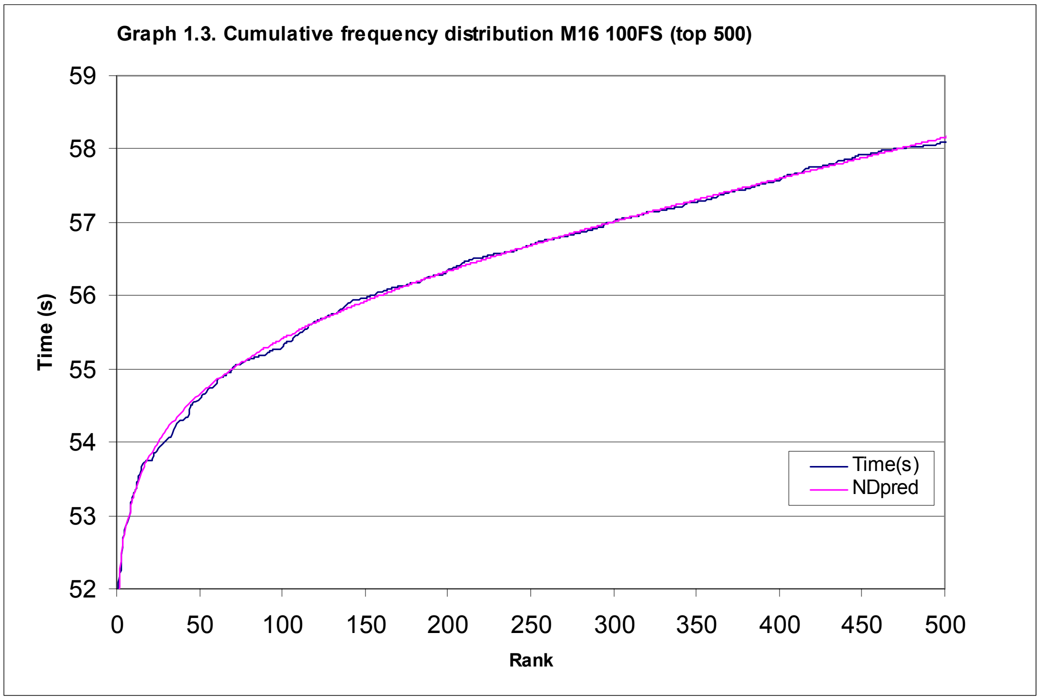


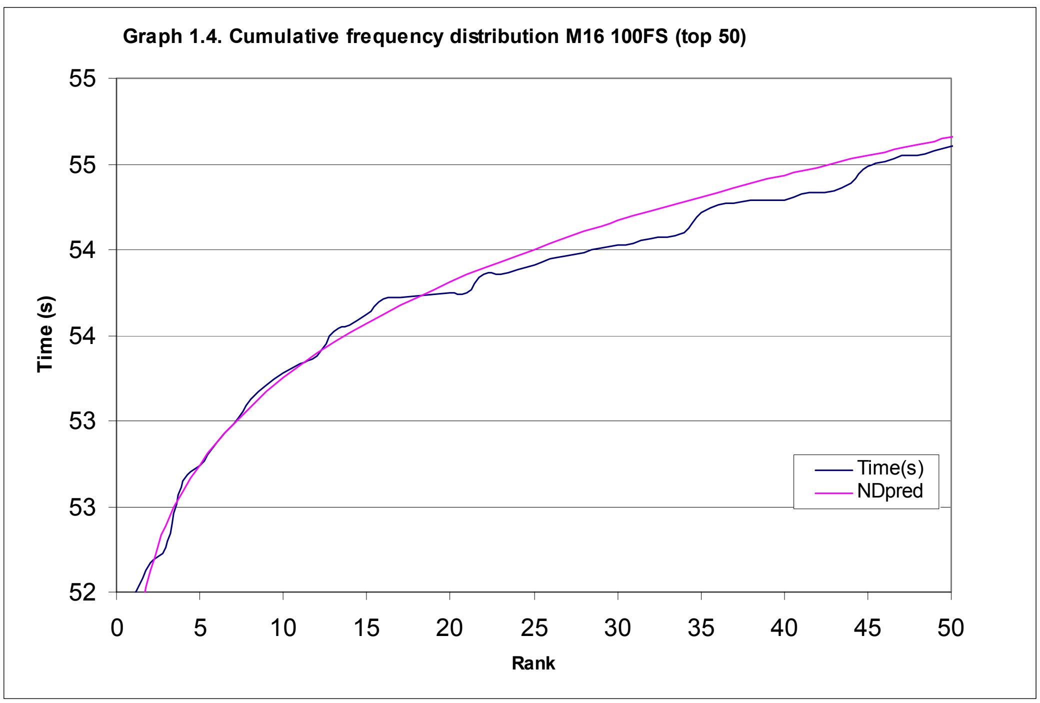


The appearance of the frequency distribution is approximately normal, but with a relatively long upper tale. There is no inherent reason to suppose that the distribution should be normal. In particular, there are lower limits to the time achieved for physical and physiological reasons, whilst there is no theoretical upper limit (although the nature of competitions which contribute to the ATB lists will tend to give a partial limit on the times). It would probably be possible to derive a transformation which would make the data approximately normal but this is beyond the scope of this report.

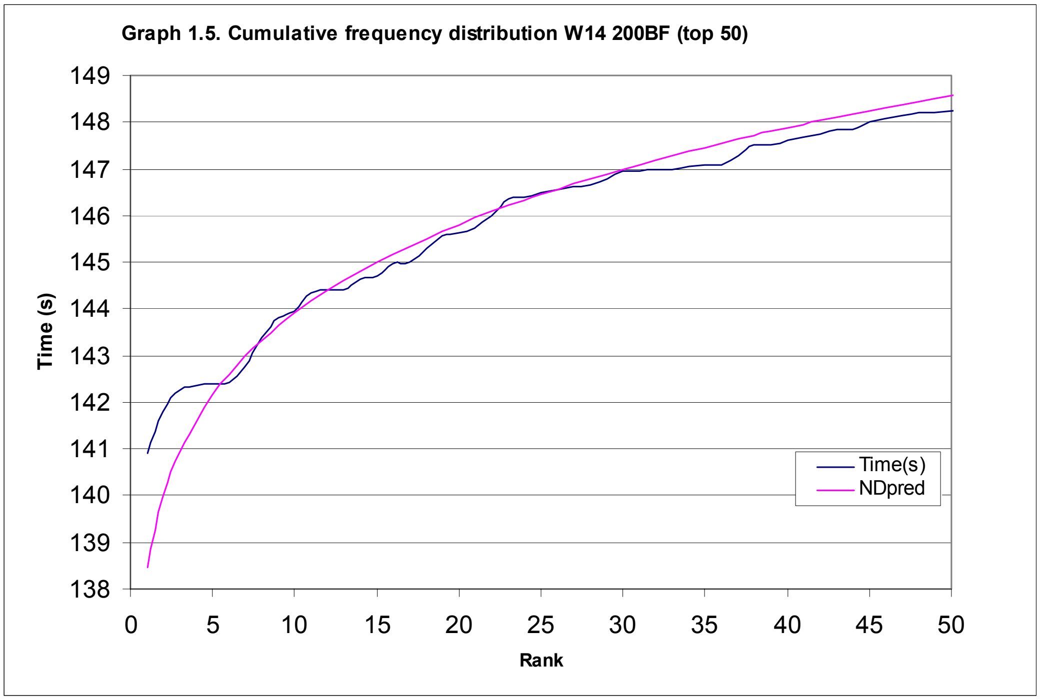
In view of the appearance of the data, and that the main interest was in the upper end, it was decided to attempt to fit a partial normal distribution to the top 25-50% of the data. The fit was carried out by assuming that the top n points of the distribution represented part of a normal distribution of m points. The values of n and m were then optimised by minimising (manually and approximately) the least square deviation of the normal fit to the first n actual points of the distribution. In the case of the event shown here the optimisation gave n=450, m=1000 leading to a distribution with a mean of 56.44 and a standard deviation of 1.27. The normal fit is shown in Graphs 1 and 2 along with the actual data. It is clear that this does not fit the bulk of the distribution, but this form of fit gives the best behaviour at the upper end.

Graph 1.3 shows the cumulative distribution compared with the fitted normal distribution for the first 500 in the ranking list. This gives a smooth curve and the fit is quite good over the entire range. The fit for the top 50 is shown in Graph 1.4. This once again is smoother than the data and the fit is seen to be not quite so good. The main point here is that only a single reference value is needed to create the qualifying times. If we look at the average of points 15 to 25 in the actual data, this is 53.78, whilst the equivalent average for the normal distribution fit is 53.80. It is, therefore, not worth using the more complicated mathematics and programming necessary for the normal distribution fit.





The equivalent graph for the women's 14 years 200 butterfly is shown as Graph 1.5. This shows the value of taking the reference value as the average of ranks 15-25 as opposed to using the top 10 where there is considerable variation in the line. The smoothness of this graph is already markedly improved over that drawn using the 2003 data and this will get even better with time.



Graph 1.5. Cumulative frequency distribution W14 200BF (top 50)
Time (s)
149
148
147
146
145
144
143
142
141
140
139
138
0
5
10
15
20
25
30
35
40
45
50
Rank
Time(s)
NDpred