Volume 38 (2), pp. 95–105
http://orion.journals.ac.za

ORiON
ISSN 0259–191X (print)
ISSN 2224–0004 (online)

A mathematical model to select the optimal age group classified freestyle relay teams for a masters swimming competition

I Fabris-Rotelli*, GD Sharp†, A Tudhope*, PJ van Staden* and M Venter‡

Received: 29 July 2022; Revised: 28 November 2022; Accepted: 28 November 2022

Abstract

This study extends the research frontier of relay swim team selection. In particular, the extension is to select the optimum teams for a club competing at a masters competition. The model uses an integer optimisation routine to select several teams of different cumulative ages who fall into the age classification categories defined at masters level. Data analysed to demonstrate the method is based on the best times of masters competitors for a swimming club in South Africa.

Key words: Integer optimisation, swimming, team selection.

1 Introduction

Swimming is a competitive sport as evidenced by events such as the quadrennial Olympic Games and biennial World Championships. Research into swimming is extensive with recent areas investigating participation uptake of elders (see [18] and [11]) to relay team strategies (see [14] and [19]) whilst traditional focus has emphasised research into athletic performance. The Journal of Swimming Research subsumed by the journal American Swimming provides ample evidence of traditional research into training methods and exercise routines for athletic performance of both elite and non-elite athletes.

The interest in competitive swimming has trickled down to club level which provides opportunities for amateur swimmers to compete and to enjoy the health benefits of regular

*Department of Statistics, University of Pretoria
†Department of Statistics, Nelson Mandela University, email: gary.sharp@mandela.ac.za
‡Department of Actuarial Science, University of Pretoria

http://dx.doi.org/10.5784/38-2-761

swimming. In addition, the benefits of exercise and competition has seen an increase in the number of participants from the older generations who have chosen to compete in organised sporting activities. Competitive swimming for the older generations is restricted to participants by age categories and fall under the label of masters competitors. This study proposes a model to allow a club to optimise the selection of several relay teams competing in a masters swimming competition.

Medic *et al.* [11] provides background into the growth of masters sports and focusses on the historical context of masters swimming. Although the primary aim of their study is to identify participation uptake within the age categories, the results show that the number of masters participants is increasing. These findings support the results of [18] who found that participation numbers of masters swimmers had increased in FINA supported competitions over a 28-year period. This interest in masters swimming provides the opportunity for further research; hence this undertaking.

2 Literature review

Team selection algorithms are not uncommon. Several have been proposed over the years for various sporting codes. Earlier research into cricket team selection by [9] and [17] garnered considerable attention sparked arguably by the phenomenal interest in the annual Indian Premier League (IPL) T20 cricket tournaments. Lemmer [10] and Bhattacharjee and Saikia [5] extend the initial research using alternative performance measurements whilst [6], [13], [2] and [1] broaden the research codes to football, fantasy team selections and athletics. Bhattacharjee and Bhattacharjee [4] compare several optimisation algorithms and conclude that the integer optimisation routine is best for their fantasy team selection scenario. An examination of the literature yielded little evidence of publications using integer optimisation to select the best swimming relay team. Brimberg *et al.* [7] discuss optimisation of a mixed medley relay team as a variation of the assignment problem whilst more closely aligned to the current research, [12] propose an integer optimisation routine for selecting a single mixed medley relay team for a 4 x 100m swimming event.

Primary research into relay team selection has targeted two areas of exploitation, the order in which the swimmers participate and the dive start of the swimmers. Saavedra *et al.* [16] evaluated relay performance and concluded that exchange block time was an important factor in the time it takes to complete the race. Fischer *et al.* [8] extended this finding by considering take off strategies defined as offensive and conservative and concluded that the conservative approach to maximise horizontal peak force was preferred. Alongside this research, [14] and [19] found evidence that team performance is influenced by the order in which the team participates. In general, the findings supported the conclusion that the fastest swimmers should swim earlier rather than later in the relay. Within the areas of swim order and diving mechanics, age and gender were found to be important factors in determining the performance of the team.

This research links the two research areas, integer optimisation and relay swimming team selection as done by [12] and extends the modelling framework to selecting multiple teams in masters swimming competitions by taking into account the effects of age and gender on team performance.

3 Methodology

Masters swimming is a competitive swimming class for older swimmers. Relay events see teams compete in yearly age groups determined by the sum of ages of the four team members. Age group categories for masters swimming relay teams start at [100 – 119] and thereafter in 40-year increments, [120 – 159], [160 – 199], [200 – 239], [240 – 279], [280 – 319], [320 – 359] and [360 – 399]. In South Africa, a pre-masters group [76 – 99] is also used. In practice, the last category is seldom able to field a team in which case the category [320 – 399] can be used. The ages of individual swimmers can vary considerably for a given cumulative age for a team. As an example, swimmers aged 24, 35, 40 and 60 can form a relay team which can compete in the age category [120 – 159]. It may seem unusual to find a 25- and 60-year-old competitor in the same group, however the combined ages of the team satisfy the age group criterion and are therefore allowable. Furthermore, such a selection may be to the advantage of the club where a younger, faster swimmer could be included in a team with a higher cumulative age. In a club with fewer members, this could also be the difference between fielding a team or not.

In South Africa, swimming competitions at masters level have a point scoring system for the swimmers who finish in the top ten places of their age category. A relay victory scores the club 20 points, a 2nd place finish scores 18 points with each minor placing decreasing the score by 2 units per decreasing order through to the 10th placing scoring 2 points. Relay races score double the points of individual races emphasising the value of selecting the best relay teams. This research proposes a modelling framework for clubs to select the optimal freestyle relay teams for a masters swimming competition to secure the highest number of points on offer for the event.

3.1 Mathematical model for freestyle relay teams

The primary objective of a club is to score the maximum points on offer from a competition. Consider the scenario where a club has two competitive teams and two age categories. If team one is exceptionally strong and team two is of average strength, the best strategy may be to rearrange the teams so that both teams are strong. This may be possible by interchanging swimmers in the teams provided their cumulative ages remain within the categories. The optimal choice resulting from this flexibility in team selection is not straight-forward and is easiest achieved computationally when comparing the cumulative time of the team against some fixed criterion. The ideal comparative criterion would be the opposing teams' times, unfortunately these are unavailable as the composition of the opposing teams is usually unknown until just prior to the start of the race. As an alternative, the national age group category records provide a proxy to compare the expected times of club relay teams. By minimising the collective times of the teams selected against the national age group records, the club would be able to select the teams per age group category most likely to perform well at the event thereby earning the most points possible.

To demonstrate this approach, the mathematical model to select the best freestyle relay teams for as many categories as possible is shown.

The binary decision variables are defined as follows; x_{ijk} is the indicator variable indicating that swimmer i is selected for age group j and gender k, for $i = 1, \ldots, n$, $j = 1, \ldots, m$ and $k = 1$ indicating males and $k = 2$ indicating females, where n is the number of swimmers available for selection and m is the number of age group categories for which selection is required. This model is designed to select a maximum of one team per age group category. If an event can have more than one team per age group category, the solution is to use this model to select the first team for as many age group categories as possible. The first set of teams will be identified, and the selected participants are then removed from the database. The unselected participants are then available for selection and the modelling process is repeated for those swimmers remaining in the database.

Using these decision variables, the objective function is defined as

$$\text{Minimise Time} = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} x_{ijk} t_{ik} - r_{jk} \right),$$

with the coefficients of objective function defined as follows. Let r_{jk} denote the national record for age group j for gender k. Let t_{ik} denote best freestyle time for swimmer i for gender k.

This objective is restricted to ensure that a four-person relay team is selected, that the cumulative ages of the four swimmers fall within the age categories and that no swimmer is selected for more than one team. These restrictions are the constraints of the model and are defined for three examples, an all-male team, an all-female team, and an equally balanced mixed relay team.

3.2 Example One: Male Freestyle Relay (set $k = 1$)

Let a_i denote the age of swimmer i. Let L_j denote the lower limit for age group j. Let U_j denote the upper limit for age group j. The constraints

$$\sum_{i=1}^{n} a_i x_{ij1} \geq L_j \quad \forall j$$

and

$$\sum_{i=1}^{n} a_i x_{ij1} \leq U_j \quad \forall j$$

ensure that the four swimmers selected for a relay team have cumulative ages that fall within the lower and upper bound of the age categories.

The constraint

$$\sum_{j=1}^{m} x_{ij1} \leq 1 \quad \forall i$$

ensures that a male swimmer is selected for a maximum of one team. This is a requirement at masters competitions to ensure that a swimmer participates only in a single relay.

To ensure that an age group team comprises of four swimmers, the constraint

$$\sum_{i=1}^{n} x_{ij1} = 4$$

is required in the optimisation model. This constraint has the potential to reduce the model to an infeasible problem. To overcome this limitation, before programming the model, an analyst should check to see if there are sufficient age appropriate participants to meet the age group requirements. In the case where there are insufficient participants, for an age group, the constraint limit for that category can be set to zero. This model will select the optimum teams per age group categories based on minimising the collective times selected against the national age group records, providing the club with the best opportunity to maximise the points.

To illustrate how this model is applied, data of masters swimmers from a club in Gauteng, South Africa is used. The data in Table 1 lists the age of the club's male swimmers and their best recent 50m freestyle time recorded in seconds. There are 32 eligible male swimmers. The objective of the club is to select the best teams per age categories from the available swimmers.

Table 1: Male swimmer times in seconds.

Nr	Age (y)	Time (s)	Nr	Age (y)	Time (s)
1	26	28	17	61	33
2	29	26	18	64	34
3	38	29	19	65	39
4	42	27	20	65	28
5	43	31	21	66	35
6	44	31	22	67	49
7	45	27	23	68	35
8	46	29	24	69	39
9	47	38	25	69	45
10	47	31	26	72	37
11	48	27	27	74	35
12	49	26	28	76	43
13	50	31	29	80	60
14	54	35	30	82	57
15	56	36	31	84	73
16	61	31	32	90	70

3.2.1 Example One: Results

To analyse the model, an integer optimisation routine is required. The solution for this data was obtained using the Microsoft Excel add-in program Solver, which can find solutions for problems with 200 integer variables or fewer. The results were confirmed in R using the package lpSolve (see [3]).

3.2.2 Discussion of results

As the first two age group categories were unable to select a team, it was necessary to include the constraints $\sum_{i=1}^{n} x_{ij1} = 0$ for $j = 1$ and 2 to ensure a feasible solution was obtained. The optimal solution yielded six teams for six different categories whilst the

Table 2: Results of freestyle relay selection for males.

Age category	Swimmer no.	Cum. Ages (y)	Cum. Times (s)	Record times (s)
[76 – 99]	nil			105.59
[100 – 119]	nil			103.80
[120 – 159]	2,3,5,10	157	117	99.04
[160 – 199]	1, 4, 6, 11	160	113	102,90
[200 – 239]	7, 8, 12, 16	201	113	101,28
[240 – 279]	13, 17, 18, 20	240	126	109,16
[280 – 319]	23, 24, 25, 27	280	154	132,53
[320 – 359]	26, 28, 30, 32	320	207	240,09

categories [76–99] and [100–119] were unable to field a team as the cumulative ages of the four youngest swimmers exceeded the upper bounds of both categories. In total the model selected 24 swimmers for the relay teams. Eight swimmers were not selected for the initial teams and were available for a second team selection.

The results obtained using a mixed-integer linear programming (MILP) algorithm are not always unique, alternative optimal solutions are often possible. Consider the two teams selected for categories [120–159] and [160–199] with swimmers 2, 3, 5 and 10 with a cumulative time of 117 seconds and swimmers 1, 4, 6 and 11 with a cumulative time of 113 seconds. An alternative solution would be to swop the teams for swimmers 1 and 2, with resulting times of 119 seconds (first category) and 111 seconds (second category). Both age categories would be legitimate thereby allowing the swop to occur. In this case, the overall objective function value does not change, the differences of the teams' swim times to the age group record change by +2 and -2 equating to no change overall. Alternative optimal solutions provide participants with some flexibility to allow friends to swim together, however the method proposed is designed to select optimal teams for the entire masters event across several age group categories.

3.3 Example Two: Female Freestyle Relay (set $k = 2$)

The same model is used to select the optimal female freestyle relay teams. The constraints and objective function remain the same albeit that the model uses the database with information of female swimmers. The data in Table 3 lists the age of the club's female swimmers and their best recent 50m freestyle time recorded in seconds. There are 36 eligible female swimmers.

3.3.1 Example Two: Results

As for example one, the solution for this data was obtained using the Microsoft Excel add-in program Solver, and the results were confirmed in R using the package lpSolve (see [3]). The results are summarised in Table 4.

Table 3: Female swimmer times in seconds.

Nr	Age (y)	Time (s)	Nr	Age (y)	Time (s)	Nr	Age (y)	Time (s)
1	21	37	13	48	30	25	56	53
2	34	31	14	48	31	26	57	38
3	35	35	15	49	37	27	58	40
4	38	30	16	50	31	28	60	63
5	38	40	17	50	70	29	63	53
6	41	29	18	51	34	30	64	42
7	41	33	19	53	32	31	65	46
8	42	32	20	54	35	32	67	45
9	44	45	21	54	35	33	69	46
10	46	37	22	55	32	34	76	55
11	46	47	23	55	48	35	81	55
12	47	48	24	55	48	36	82	78

Table 4: Results of freestyle relay selection for females.

Age category	Swimmer no.	Cum. Ages (y)	Cum. Times (s)	Record times (s)
[76 – 99]	nil			132,04
[100 – 119]	nil			120,79
[120 – 159]	2, 3, 4, 8	149	128	112,58
[160 – 199]	6, 7, 14, 15	179	130	117,63
[200 – 239]	13, 16, 18, 19	202	127	116,04
[240 – 279]	21, 22, 30, 32	240	154	132,99
[280 – 319]	20, 33, 34, 35	280	191	202,2
[320 – 359]	nil			333,75

3.3.2 Discussion of results

Three age group categories were unable to field a team. The youngest team possible with cumulative age of 128 excluded the first two categories, whilst the oldest team possible with cumulative age of 308 excluded the last category [320 – 359]. The results in Table 4 show that five age group teams comprising 20 female swimmers were selected. The swimmers selected for the relay teams are listed in Table 4, with expected relay times and cumulative ages. An interesting observation from the results is that as the category ages increase, the cumulative ages of the team are closer to the lower bound, this is not necessarily the case for the younger age group categories. The reason for this is that the model is able to ensure age limits are satisfied by combining a potentially faster, younger swimmer with a group of older swimmers in the older categories. In addition, these results lend support to the findings of [11] that age is an important factor in swimming times.

3.4 Example Three: Mixed Freestyle Relay

In mixed relay events the team composition requires two male and two female swimmers. This rule requires that the constraints of the model be modified to ensure a balanced team.

The constraint

$$\sum_{j=1}^{m} x_{ijk} \leq 1 \ \ \forall i \text{ and } k$$

ensures no swimmer is selected more than once. The constraints

$$\sum_{i=1}^{n} x_{ij1} = 2 \ \ \forall j$$

and

$$\sum_{i=1}^{n} x_{ij2} = 2 \ \ \forall j$$

ensure that each team includes two male and two female swimmers. In age group categories which are unable to field a team, the constraint limit is changed to zero.

As the age groups are the same for both genders the limits for either k can be used, that is the age category boundaries are set by the constraints,

$$\sum_{i=1}^{n} a_i x_{ij1} \geq L_j \ \ \forall j,$$
$$\sum_{i=1}^{n} a_i x_{ij1} \leq U_j \ \ \forall j.$$

The data in Table 5 lists the age of the club's swimmers and their best recent 50m freestyle time recorded in seconds. There are 68 eligible swimmers.

3.4.1 Example Three: Results

The results of the model are summarised in Table 6.

3.4.2 Discussion of results

Seven relay teams were possible for this dataset. Age group category [76 – 99] was unable to field a team as the youngest possible team had a cumulative age of 110. The team selection results are summarised in Table 6 with the swimmer no. identifying the participant selected for the team. It is observed that as the age group categories increase, the cumulative age of the team selected, is close to the lower bound of the category. This result agrees with prior research indicating that age is a factor in swimming performance and that older participants typically swim slower times than younger participants.

4 Conclusion

The mathematical modelling of team selection provides a simple approach to choose multiple teams simultaneously. Provided the objective function chooses a functional form to minimise swimming times subject to a valid criterion, in this case existing national

Table 5: All swimmers available for selection.

No	Age	Time	Gender	No	Age	Time	Gender	No	Age	Time	Gender
1	26	28	1	24	69	39	1	47	49	37	2
2	29	26	1	25	69	45	1	48	50	31	2
3	38	29	1	26	72	37	1	49	50	70	2
4	42	27	1	27	74	35	1	50	51	34	2
5	43	31	1	28	76	43	1	51	53	32	2
6	44	31	1	29	80	60	1	52	54	35	2
7	45	27	1	30	82	57	1	53	54	35	2
8	46	29	1	31	84	73	1	54	55	32	2
9	47	38	1	32	90	70	1	55	55	48	2
10	47	31	1	33	21	37	2	56	55	48	2
11	48	27	1	34	34	31	2	57	56	53	2
12	49	26	1	35	35	35	2	58	57	38	2
13	50	31	1	36	38	30	2	59	58	40	2
14	54	35	1	37	38	40	2	60	60	63	2
15	56	36	1	38	41	29	2	61	63	53	2
16	61	31	1	39	41	33	2	62	64	42	2
17	61	33	1	40	42	32	2	63	65	46	2
18	64	34	1	41	44	45	2	64	67	45	2
19	65	39	1	42	46	37	2	65	69	46	2
20	65	28	1	43	46	47	2	66	76	55	2
21	66	35	1	44	47	48	2	67	81	55	2
22	67	49	1	45	48	30	2	68	82	78	2
23	68	35	1	46	48	31	2				

Table 6: Results of mixed freestyle relay selection.

Age category	Swimmer no.	Cum. Ages (y)	Cum. Times (s)	Record times (s)
[76 – 99]	Nil			105,59
[100 – 119]	1, 2, 33, 38	117	120	103,80
[120 – 159]	4, 7, 34, 36	159	115	99,04
[160 – 199]	8, 12, 40, 45	185	117	102,90
[200 – 239]	11, 16, 46, 48	207	120	101,28
[240 – 279]	20, 23, 51, 54	241	127	109,16
[280 – 319]	26, 28, 62, 65	281	168	132,53
[320 – 359]	27, 32, 66, 67	321	215	240,09

records, standard integer linear program (ILP) algorithms provide solutions to the selection problem. This research extends the modelling approach of [12] to multiple team selection across independent age categories.

In the case where a club can enter more than one team per age group category the modelling process could easily be repeated for those swimmers not originally selected. In example one, eight swimmers were available for selection in the next round. These eight swimmers can then be selected for teams and enjoy the opportunity to participate at a competitive event.

In future, this research could be extended to include team selection for medley relay teams, whilst in athletics there exists an opportunity to apply this modelling framework to mixed relay team events. The 2022 athletics world championships in Eugene, Oregon demonstrated the importance of team strategies, where the 4 x 400m mixed relay team of the USA were unable to maintain their lead and placed third to the faster finishing teams of the Dominican Republic and the Netherlands.

References

[1] Bailey MD & Nowak M, 2018, *MeetOpt: A multi-event coaching decision support system*, Decision Support Systems, 112: 60–75.

[2] Becker A & Sun XA, 2016, *An analytical approach for fantasy football draft and lineup management*, Journal of Quantitative Analysis in Sports, 12(1): 17-30.

[3] Berkelaar M, 2020, *lpsolve: To solve linear/Integer Programs*, R package version 5.6.15.

[4] Bhattacharjee B & Bhattacharjee D, 2021, *Comparing performance of different team selection methods using fantasy league scoring system*, International Journal of Human Movements and Sports Science 9(6): 1352–1361.

[5] Bhattacharjee D & Saikia H, 2016, *An objective approach of balanced cricket team selection using binary integer programming method*, Opsearch 53(2): 225-247.

[6] Brettenny WJ, 2010, *Integer optimisation for the selection of a fantasy league cricket team*, MSc dissertation, Nelson Mandela University.

[7] Brimberg J, Ladany SP & Hurley WJ, 2018, *Choosing a winning team for mixed medley events*, International Journal of Operational Research 31(3): 300-312.

[8] Fischer, S, Braun C & Kibele A, 2017, *Learning relay start strategies in swimming: What feedback is best?*, European Journal of Sport Science 17(3): 257-263.

[9] Gerber H & Sharp GD, 2006, *Selecting a limited overs cricket squad using an integer programming model*, Research in Sport, Physical Education and Recreation 28(2): 81–90.

[10] Lemmer H, 2013, *Team selection after a short cricket series*, European Journal of Sport Science 13(2): 200-206.

[11] Medic N, Mussener M, Lobinger BH & Young BW, 2019, *Constituent year effect in Masters sports: An empirical view on the historical development in US Masters Swimming*, Journal of Sports Science and Medicine 18(3): 505-512.

[12] Ning P & Chen C, 2021, *Development and application of a system for optimal combination of decision in the mixed 4 x 100m medley relay*, International conference on information technology and contemporary sports. IEEE, 312–317.

[13] Ozceylan E, 2016, *A mathematical model using AHP priorities for soccer player selection: a case study*, South African Journal of Industrial Engineering 27(2): 190-205.

[14] Qiu X, Veiga S, Lorenzo A, Kibele A & Navarro E, 2021, *Differences in the key parameters of the individual versus relay swimming starts*, Sports Biomechanics Published online 16 March 2021. DOI: 10.1080/14763141.2021.1878262

[15] R Core Team, 2020, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[16] Saavedra JM, Garcia-Hermoso A, Escalante Y, Dominguez AM, Arellano R & Navarro F, 2014, *Relationship between exchange block time in swim starts and final performance in relay races in international championships*, Journal of Sports Sciences 32(19): 1783-1789.

[17] Sharp GD, Brettenny WJ, Gonsalves JW, Lourens M & Stretch RA, 2011, *Integer optimisation for the selection of a Twenty20 cricket team*, Journal of the Operational Research Society 62(9): 1688-1694.

[18] Unterweger CM, Knechtle B, Nikolaidis PT, Roseman T & Rust CA, 2016, *Increased participation and improved performance in age group backstroke master swimmers from 25–29 to 100–104 years at the FINA World Masters Championships from 1986 to 2014*, Springerplus 5(1): 1-12.

[19] Wu PPY, Babaei T, O'Shea M, Mengersen K, Drovandi C, McGibbon KE, Pyne DB, Mitchell LJG & Osborne MA, 2021, *Predicting performance in 4 x 200-m freestyle swimming relay events*, PLoS ONE 16(7): e0254538. https://doi.org/10.1371/journal.pone.0254538.