

ATACP – Good Boost

Audit review of Good Boost AI decision
making by ATACP

Report summary
June 2019



**Aquatic Therapy Association of
Chartered Physiotherapy**



good boost

Introduction

This report summarises the audit review of Good Boost's exercise recommendation for therapeutic aquatic activity by artificial intelligence (AI). Overall, Good Boost performed well against random and human therapeutic exercises selection in 10 case study scenarios. Good Boost AI was lower performing in hip patient scenario exercises compared to knee exercises and presents an area for future development.

Each case study was presented with background information and clinical information for the patient. The ATACP reviewer was presented with three different groups (A, B & C) of exercises (which were presented randomly to avoid bias). These three groups of exercises were either random, selected by a qualified therapist or Good Boost's AI decision making.

ATACP reviewers were asked to score each exercise group between 0-100 for safety, appropriateness and optimisation. The scoring system is shown below.

1. **Safety** - It is safe for the individual to participate given the information available
2. **Appropriateness** - It is appropriate/suitable set of exercises for the individual given the information available regarding condition
3. **Optimisation** – Is it an optimal set of exercises given the information on the individual / condition

The scoring guidance (5 scoring bandings in total):

0-40 - Failed recommendation

40-55 - Suitable/average recommendation

55-70 - Above average recommendation

70-85 - High value recommendation

85-100 – Outstanding value recommendation

Results

The results tables below show the average score from ATACP reviewers. Each case shows the comparison of AI to Random and AI to human below each case. Where the score for AI was greater than either random or human this is highlighted green. Cells highlighted in red display the percentage difference where scores were less than random or human. Any scores in red indicate a failed exercise selection following the review scoring system.

Case Study 1

	Safe	Appropriate	Optimal
RAN	62.5	47.5	42.5
HUM	65	55	47.5
AI	70	53.5	50
AI: Random	11%	11%	15%
AI: Human	7%	-3%	5%

Case Study 2

	Safe	Appropriate	Optimal
RAN	52.5	37.5	40
HUM	73	55.5	48
AI	62	45	45
AI: Random	15%	17%	11%
AI: Human	-18%	-23%	-7%

Case Study 3

	Safe	Appropriate	Optimal
RAN	27.5	32.5	30
HUM	52.5	42.5	35
AI	47.5	47.5	40
AI: Random	42%	32%	25%
AI: Human	-11%	11%	13%

Case Study 4

	Safe	Appropriate	Optimal
RAN	39	37.5	37.5
HUM	62.5	47.5	47.5
AI	61.5	45	41
AI: Random	37%	17%	9%
AI: Human	-2%	-6%	-16%

Case Study 5

	Safe	Appropriate	Optimal
RAN	62	47.5	49.5
HUM	67	50	45
AI	67	50	49.5
AI: Random	7%	5%	0%
AI: Human	0%	0%	0%

Case Study 6

	Safe	Appropriate	Optimal
RAN	65	50	47.5
HUM	65	52	49.5
AI	67.5	52.5	45
AI: Random	4%	5%	-6%
AI: Human	4%	1%	-10%

Case Study 7

	Safe	Appropriate	Optimal
RAN	42.5	35	35
HUM	62.5	42.5	42.5
AI	67.5	42.5	40
AI: Random	37%	18%	13%
AI: Human	7%	0%	-6%

Case Study 8

	Safe	Appropriate	Optimal
RAN	65	45	40
HUM	69.5	52	49.5
AI	37.5	35	32.5
AI: Random	-73%	-29%	-23%
AI: Human	-85%	-49%	-52%

Case Study 9

	Safe	Appropriate	Optimal
RAN	60	45	45
HUM	60	45	42.5
AI	50	40	40
AI: Random	-20%	-13%	-13%
AI: Human	-20%	-13%	-6%

Case Study 10

	Safe	Appropriate	Optimal
RAN	67.5	40	37.5
HUM	69.5	52	49.5
AI	62.5	42.5	42.5
AI: Random	-8%	6%	12%
AI: Human	-11%	-22%	-16%

Discussion

With the exception of Case Study 8, Good Boost AI scores above 40 in safety, appropriateness and optimal for all cases. The clinical characteristics in case study 8 will be reviewed following feedback from ATACP reviewers. The key areas for improvement are around recommendation for users with hip conditions. This will be included in to Good Boost's priority for improvement and development and hip cases will be included again for the next ATACP audit.

Discussion

Outcomes comparisons are the comparison of the score for Good Boost AI for a case studies exercises on safety, appropriateness and optimal scores.

There are 28 outcome comparisons between AI human/random where Good Boost AI scores higher than human or random. There are 5 comparisons where Good Boost AI scores equal to human or random and 27 comparisons between AI and human or random where Good Boost AI scores lower.

AI compared to Random performance

In each comparison between Good Boost AI and random, Good Boost outperforms random in score for 21 (70%) outcomes, equal in 1 (3%) outcome and underperform random in 8 (27%) outcomes.

AI compared to Human therapist performance

In each comparison between Good Boost AI and human, Good Boost outperforms random in score for 7 (23%) outcomes, equal in 4 (13%) outcome and underperform random in 19 (63%) outcomes.

AI performance in all outcomes that are average or above

For each exercise set including the score of safety, appropriateness and optimal, Good Boost scores 40-55 'average' in 20 outcomes (66%). Good Boost score 55-70 'above average' 6 outcomes (20%). Good Boost scores 70+ 'high performance decision' in 1 outcome (3%). Good Boost scores below 40 in 3 outcomes (9%), which are all in the same case study 8.

Conclusion

Overall, with the exception of case study 8, Good Boost scores average or above for safety, appropriateness and optimal exercises selection. Good Boost AI outperforms random selection with 70% of decisions scoring higher for each comparison outcome. AI underperforms human therapist with 23% of decision scoring higher for each comparison outcome.

It demonstrates that there is room for development and improvement for Good Boost's decision making against a human therapist. Nonetheless, Good Boost's decision making is selecting exercises which 'average' or 'above average' in quality for exercise safety, appropriateness and optimisation in 91% of decisions. This indicates that Good Boost has created the foundation to build on for the user of artificial intelligence decision making for therapeutic aquatic exercises selection.