Assessment of Treatment Response Through the Use of Personalized Endpoints: Using Artificial Intelligence to Assist Goal Attainment Scaling

Gunes Sevinc[1], Katie Crespo[1], Andrea Escoto[1], Kari Pope[1], Chere Chapman[1], Susan E Howlett[1,2,3], Kenneth Rockwood[1,3,4]

[1]Ardea Outcomes, Halifax, NS, Canada.
[2]Department of Pharmacology, Dalhousie University, Halifax, NS, Canada.
[3]Geriatric Medicine Research Unit, Nova Scotia Health Authority, Halifax, NS, Canada.
[4]Division of Geriatric Medicine, Dalhousie University, Halifax, NS, Canada.

Outline

What is Goal Attainment Scaling?

When should we use AI to assist GAS?

How should we use AI?

How does “Goal Attainment Scaling” work?

An Overview

Goal Setting

1 Identify Goals
The clinician facilitates the interview with the patient and/or caregiver to identify 3-5 goals.

2 Build GAS Scales
Together they develop a 5-point goal attainment scale for each identified goal.

Follow-up: Scoring Goal Attainment

3 Obtain Current Status
The patient and clinician discuss the patient’s current status concerning each goal area.

4 Assess Goal Attainment Level
The patient and clinician rate the level of attainment for each goal.

Goal Scale		*Goal Area: <Goal Area Title>*
Attainment Levels		
+2	Much more than expected	
+1	Somewhat more than expected	
0	Expected Outcome*	
-1	Somewhat less than expected	
-2	Much less than expected	

*Attainment level that can be realistically achieved by the end of the designated assessment period.

How does “Goal Attainment Scaling” work?

Goal-setting visit

What matters most to you?

Goal Title	
Much Better than the Goal	+2
Somewhat Better than the Goal	+1
The Goal	0
Baseline Status	-1
Much Worse than the Goal	-2

How does “Goal Attainment Scaling” work?

Goal-setting visit

Scenario

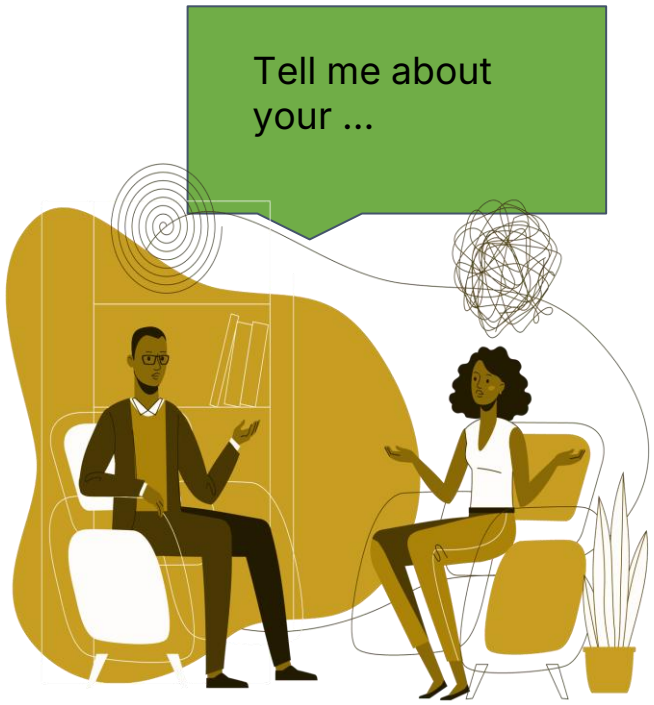
M is a 6-year-old boy diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD). His caregivers fear for his safety as he tends to be very impulsive and do dangerous things, such as running out of the house and into the street without checking. His caregiver has tried teaching him to wait for permission before approaching the street. Unfortunately, this continues to be an issue, and M runs into the street without permission or supervision 4-6 times per week.

Disease Area: ADHD Goal Area: Impulsivity	
+2	Much better than expected The participant does not run into the street without his caregiver’s permission or supervision.
+1	Somewhat better than expected The participant runs into the street only once per week.
0	Expected level of outcome/goal level The participant runs into the street 2-3 times per week.
-1	Somewhat less than expected/baseline level The participant runs into the street without his caregiver's permission or supervision 4-6 times per week.
-2	Much less than expected The participant runs into the street 7 or more times per week.

How does “Goal Attainment Scaling” work?

Follow-up visit(s)

	Participant Rating	GAS Interviewer Rating
Much Better than the Goal	○ +2	○ +2
Somewhat Better than the Goal	○ +1	○ +1
The Goal	○ 0	○ 0
Baseline Status	○ -1	○ -1
Much Worse than the Goal	○ -2	○ -2



How does “Goal Attainment Scaling” work?

The scoring

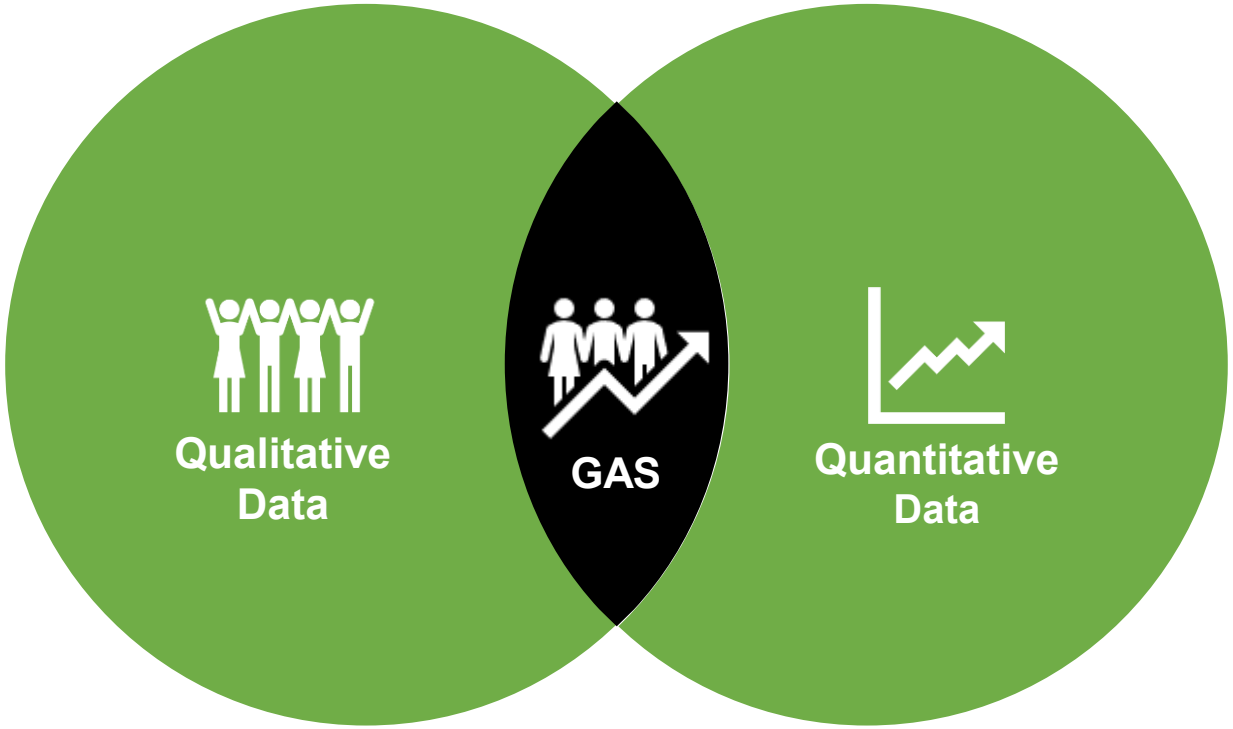
Participants are encouraged to set 3-5 goals

Goals can be **ranked** by difficulty or importance

Formula corrects for differential weighting and number of goals between patients

*For each participant, a **summary score** reveals the extent to which goals have been attained*

GAS data provides multiple layers of information



Where quantitative and qualitative meet to measure efficacy and effectiveness

Why should we assess treatment response using personalized outcome assessments?

Highly responsive

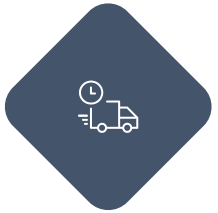
No dilution of the treatment effect due to mixing affected and unaffected patients

Inherently clinically meaningful and nonarbitrary

Clinician and patients set goal considering clinical meaningfulness.

Reveals what's important to patients

No need for justification of the measure for regulatory purposes



Minimizes recall and other biases

The scale consists of five levels that are specific lowering the possibility of recall and other biases.

The problem: Psychometric properties of the scales. Can we leverage AI?

Standardization

Minimize variability through training and other tools to ensure psychometric adequacy of the goal scales

Goal relevance

Setting goals, they are relevant to the treatment & are likely to be affected by it through goal inventories

AI can assist:

Standardizing the GAS process through clinician training and utilizing artificial intelligence (AI) to complement formal training may help minimize inconsistencies and improve the robustness of goal scales.

How?

What is the appropriate prompting strategy?

Automated Quality Checks: Natural language processing (NLP) can assess goal phrasing for adherence to psychometric criteria and detect ambiguities.

Define what to check. Identify core quality criteria.

Compare prompting strategies.

Establish goal standard, compare AI review with expert reviewed benchmarks.

Evaluate performance, assess reliability and accuracy.

Refine and test usability.

Step 2: Prompt types to guide the AI's reasoning:

- **Batch prompting:** Input multiple goals with minimal context for rapid screening.
- **Contextual prompting:** Add clinical background or patient descriptors to test whether context improves goal evaluation.
- **Chain-of-thought prompting:** Ask the AI to reason stepwise ("Explain why this goal is or isn't measurable").
- **Multi-turn prompting:** Simulate interactive refinement ("How could this goal be more specific?").

Testing prompting strategies for AI-assisted goal setting:

Developed goal scales (n=4) with common errors:

4

- Overlapping attainment levels
- Vagueness
- Multidimensionality
- Incorrect ordering of levels

Evaluated the goal scales using predefined psychometric criteria (n=20) using the following strategies:

20

- Simple batch prompting
- Batch prompting with context and rationale for each goal-scale criteria
- Incorporating chain-of-thought prompting into the previous prompt
- Multi-turn prompting with context and rationale

Results: Multi-turn prompting with context and rationale

Goal Scale	Error(s) in Goal Scale	Number of times error was detected in three trials (n, %)*			
		Batch prompting	Batch prompting with context and rationale	Chain-of-thought prompting	Multi-turn prompting with context and rationale
1	Incorrect ordering of attainment levels	0/3 (0%)	0/3 (0%)	0/3 (0%)	2/3 (67%)
2	Overlap	1/3 (33%)	1/3 (33%)	3/3 (100%)	3/3 (100%)
3	Multidimensionality	3/3 (100%)	3/3 (100%)	3/3 (100%)	3/3 (100%)
4	Vagueness	2/3 (67%)	3/3 (100%)	3/3 (100%)	3/3 (100%)
	SUCCESS RATE:	50%	58.25%	75%	91.75%

*Four prompting strategies were used. Prompt 1 involved simple batch prompting. Prompt 2 included batch prompting with additional context and rationale. Prompt 3 integrated chain-of-thought prompting into the previous approach. Finally, Prompt 4 employed multi-turn prompting, also with context and rationale.

Conclusion & Discussion

While some inconsistencies in error detection persisted, chain-of-thought and multi-turn prompting prompting enabled AI to provide feedback that could enhance goal scales.

AI-assisted feedback may support GAS training and monitoring and, ultimately, may be utilized to strengthen the psychometric properties of individual goal scales.

Potential use: AI-assisted training to the clinician while they are developing the scales with the patient

Potential impact: Low Risk High Benefit

Acknowledgements

The National Research Council of Canada
Industrial Research Assistance Program

LinkedIn
ca.linkedin.com/company/ardeaoutcomes

X
@ardeaoutcomes

Email
gunes.sevinc@ardeaoutcomes.com

Phone
+1(902) 410-3888