Non-Inferiority of an Ensemble AI Model Versus Physician-Guided ESA Dosing in Hemodialysis

Retrospective Development and Randomized Controlled Trial

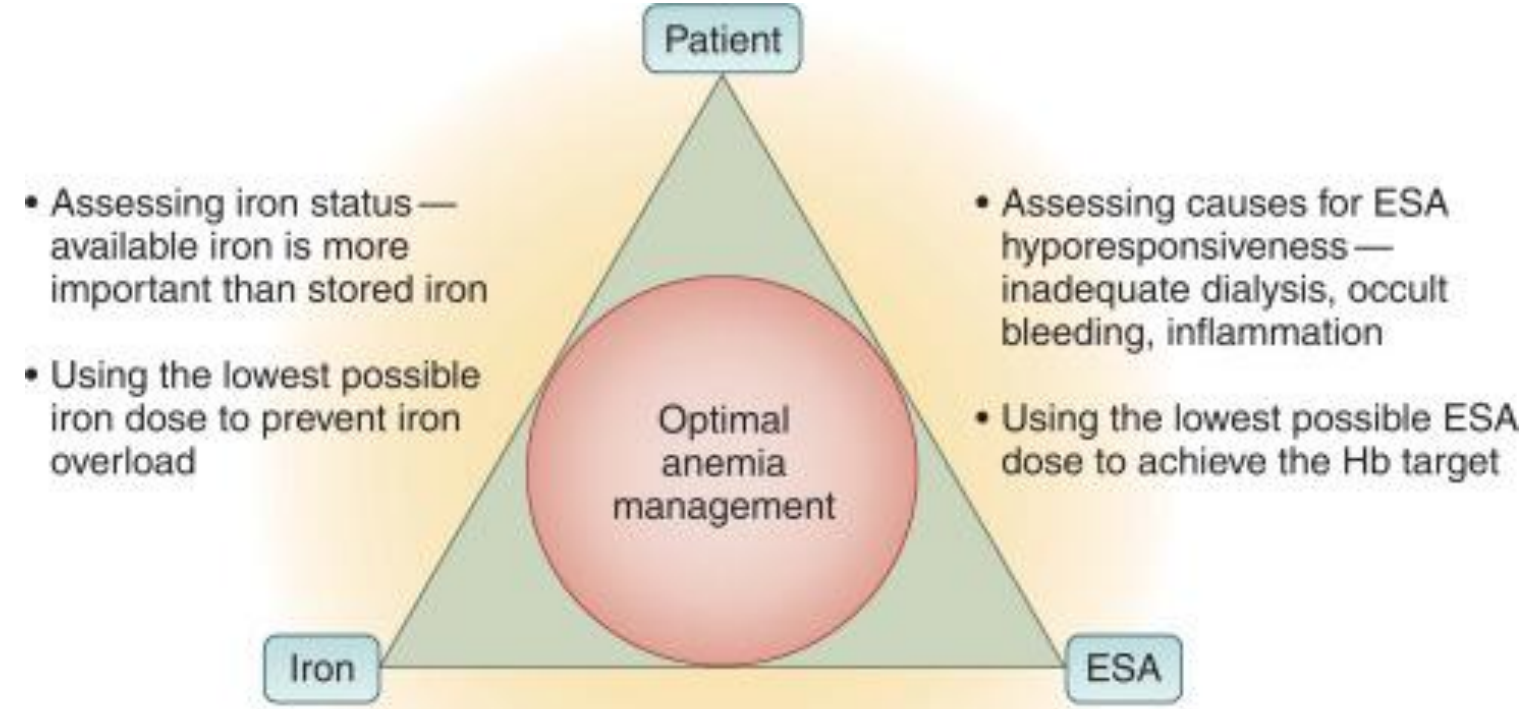
Ping-Hsun Wu, Associate Professor

Yi-Wen Chiu, Chan Hsu, Min-Yen Lin, and Yihuang, Kang

Division of Nephrology, Department of Internal Medicine
Kaohsiung Medical University Hospital

2025.12.05

Interaction of patient, erythropoiesis-stimulating agent (ESA), and iron in the management of anemia in chronic kidney disease



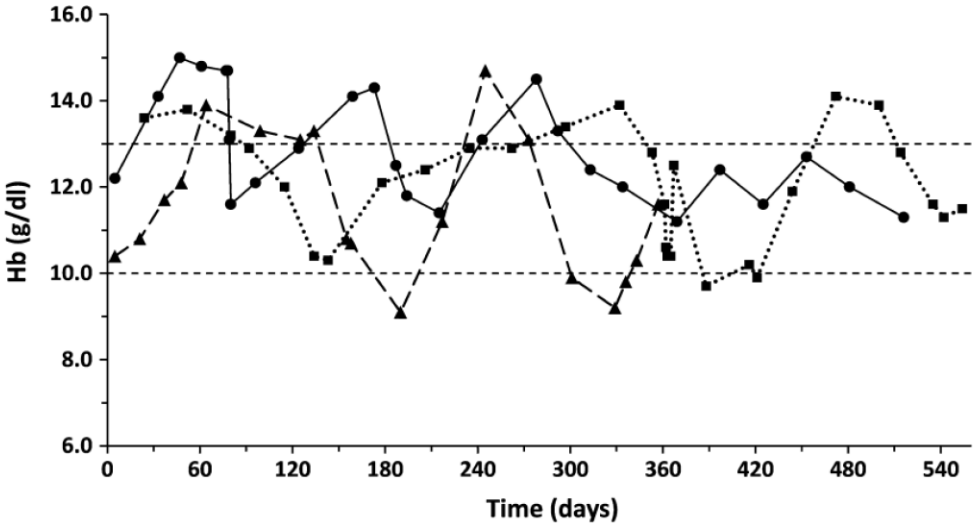
Kidney Int. 2014 Oct;86(4):676-8

KDIGO guideline 2025

- The target Hb range is **10–11.5 g/dL**
- **Higher targets** are associated with **increased risk of stroke, hypertension, and cardiovascular events**
  - **Epoetin alfa or beta**: 20–50 IU/kg three times per week, administered intravenously (IV) or subcutaneously (SC). **Eprex, Recormon**
  - **Darbepoetin alfa**: 0.45 µg/kg once weekly (IV or SC), or 0.75 µg/kg once every two weeks (SC). **Aranesp**
  - Continuous erythropoietin receptor activator (**CERA**, e.g., methoxy polyethylene glycol-epoetin beta): 0.6 µg/kg every two weeks (IV or SC). **Mircera**

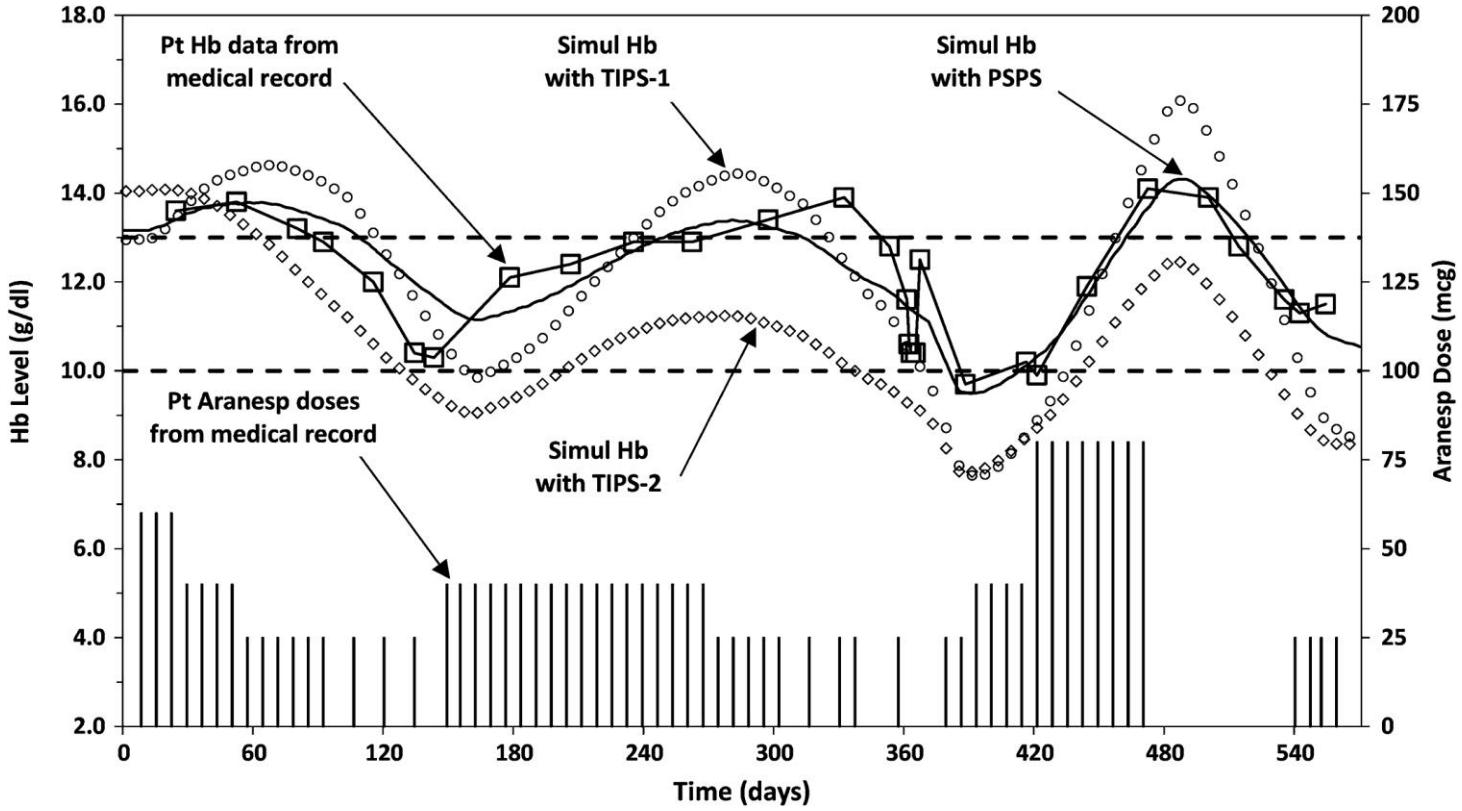
Potential mechanism of increased cardiovascular risk with higher hemoglobin targets in ESA studies

Kidney Int Suppl (2011). 2017 Dec;7(3):157-163



Titration of ESA Dosage: Given IV once weekly (maximum dose 300 mcg/week)	
Hb < 10 g/dl	Increase by 3 vials
Hb 10.1–10.5 g/dl	Increase by 2 vials
Hb 10.6–11.4 g/dl	Increase by 1 vial
Hb 11.5–12.5 g/dl	No change in dose
Hb 12.6–13.9 g/dl	Decrease by 1 vial
Hb ≥ 14 g/dl	HOLD 2 weeks; then decrease by 2 vials

Available vial sizes (mcg) 25, 40, 60, 100, 150, 200, 300.



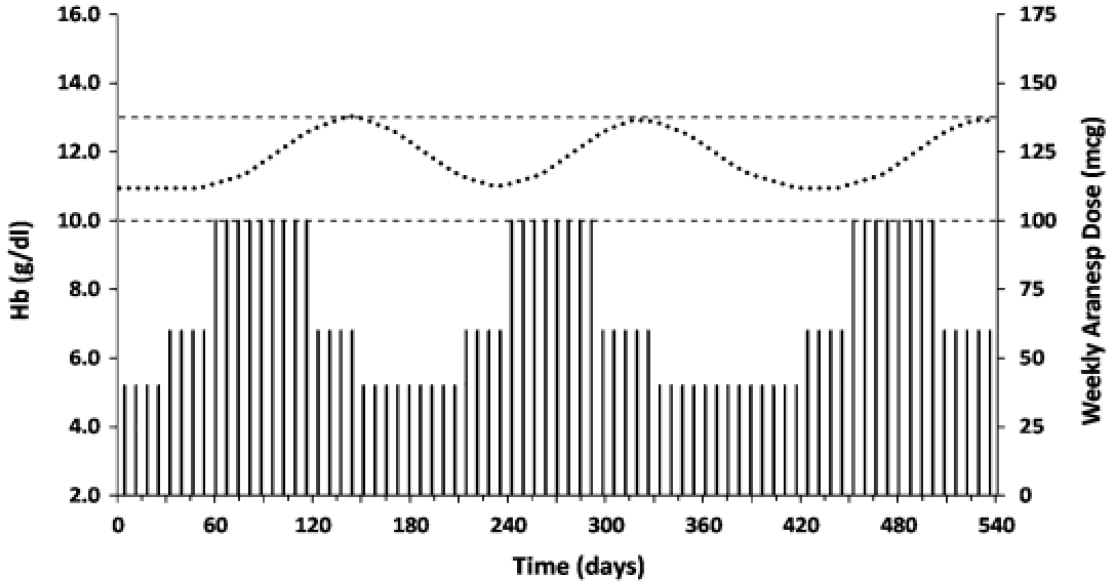
Representative Monte Carlo **simulations** with tentative **individual parameter sets (TIPS)** to identify the **best-fit** (MSE) **patient-specific parameter set (PSPS)**.

MAYO CLINIC

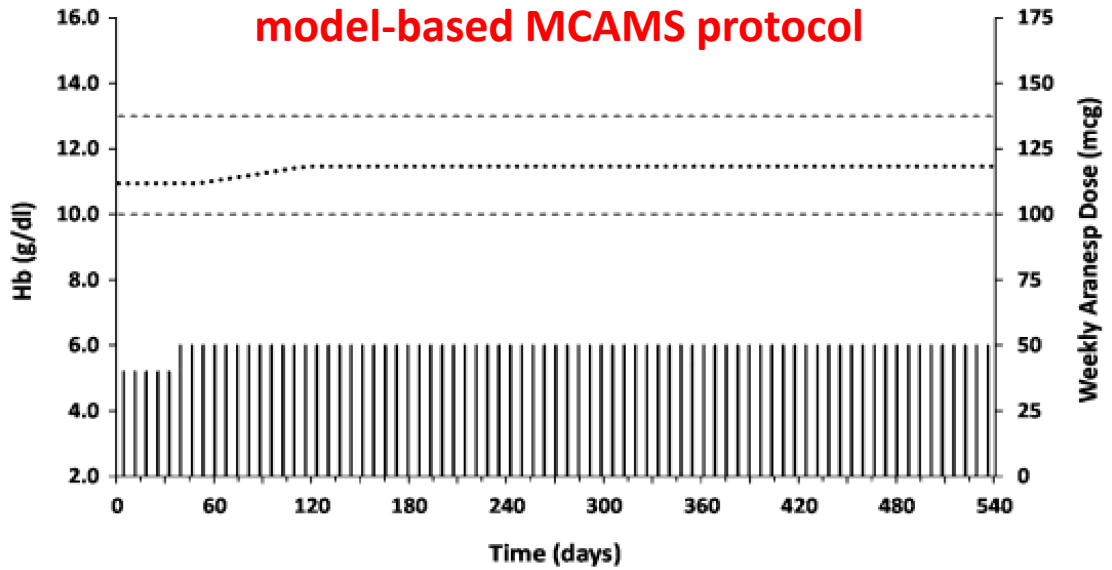
Semin Nephrol. 2018 Jul;38(4):410-417

The Mayo Clinic Anemia Management System (MCAMS)

Standard ESA protocol

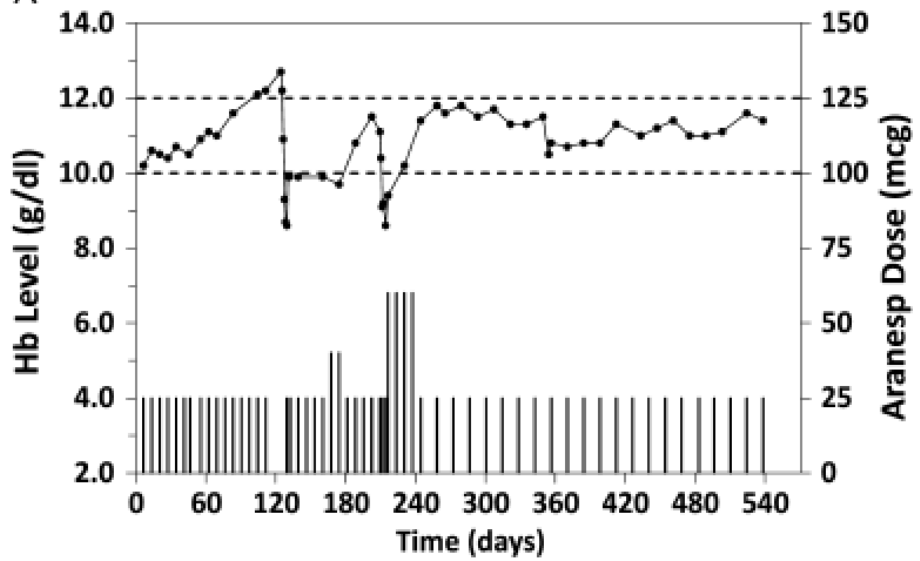


model-based MCAMS protocol

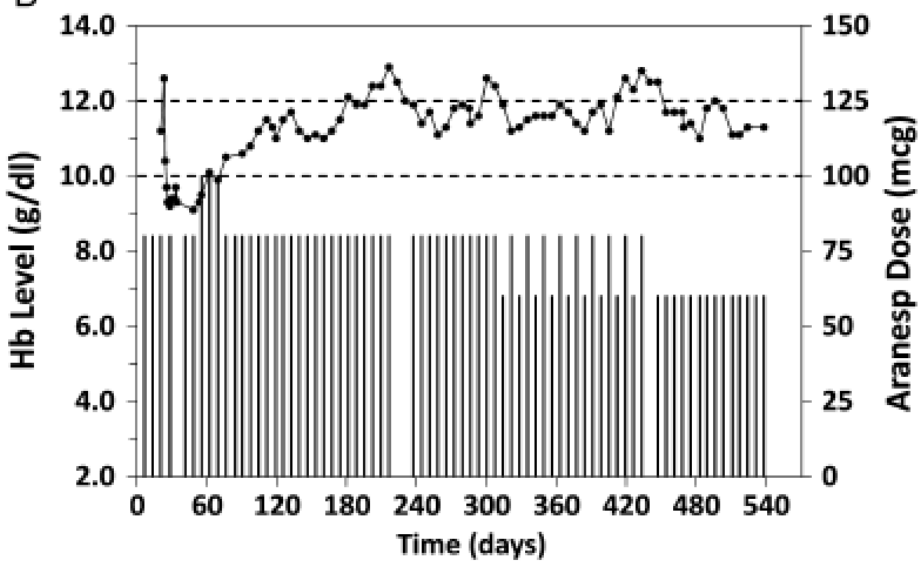


Therapeutic outcomes: clinical Hb results in ESRD patients following **model-based MCAMS decision rules** to identify and adjust ESA dosage regimens

A



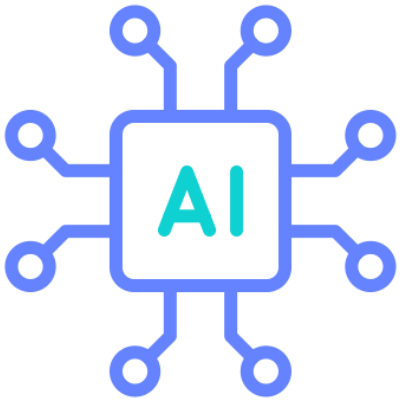
B



Semin Nephrol. 2018 Jul;38(4):410-417

Research questions

The challenge of optimizing ESA dosing in HD patients



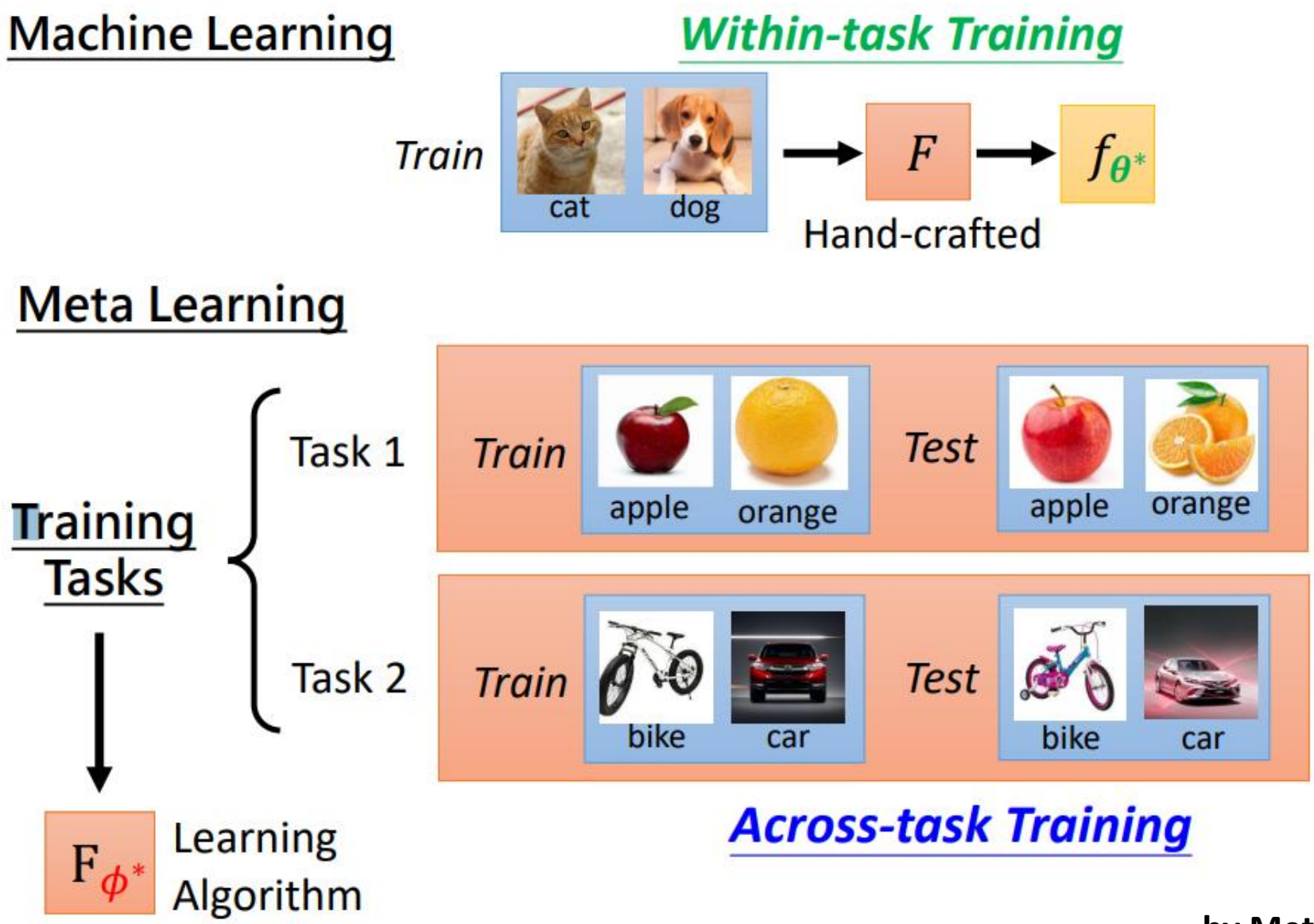
V.S.

AI-guided group (Meta-learning model)

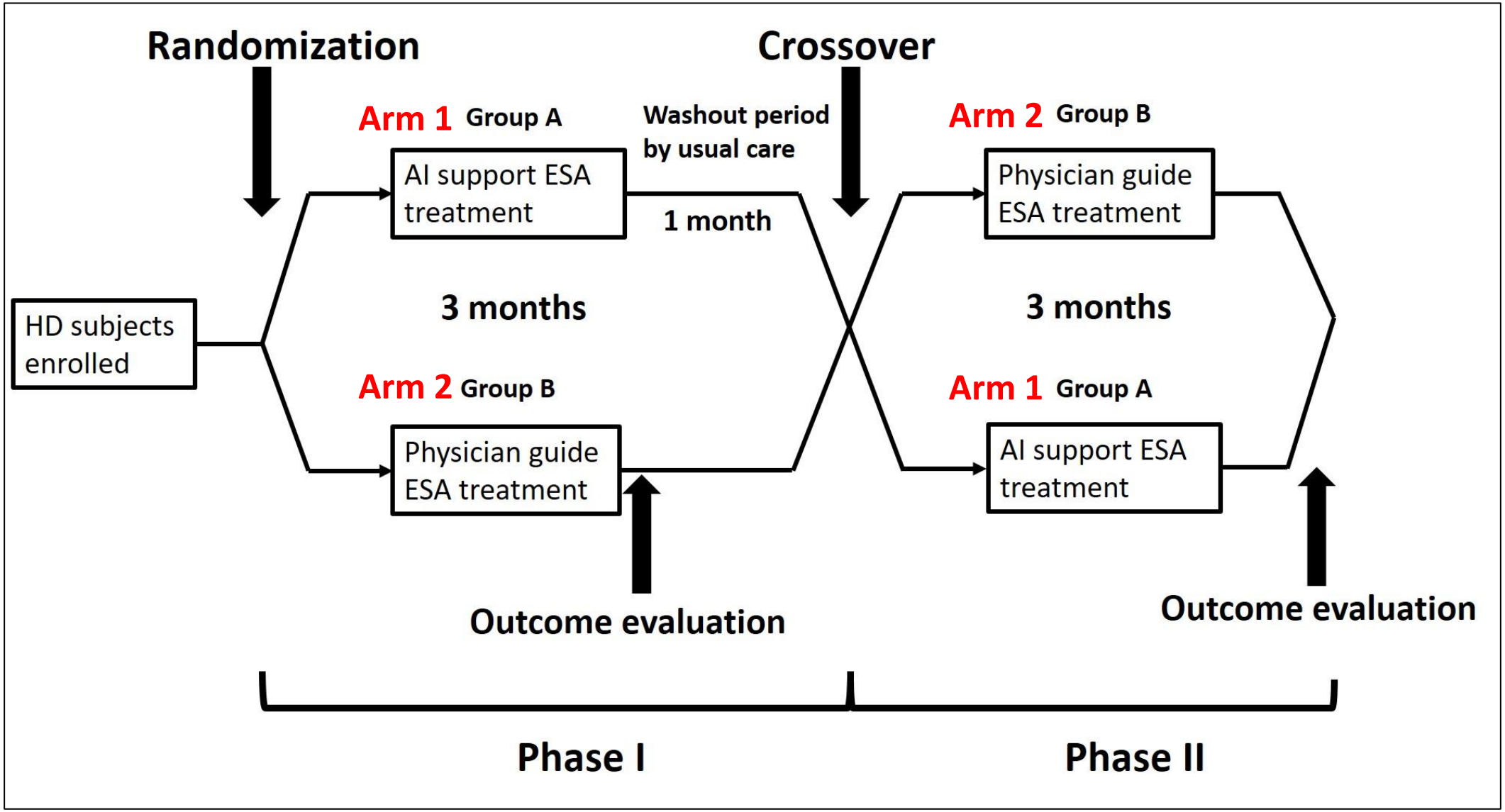
Meta-learning models are **"learning-to-learn" systems** that are trained across many related tasks

Physician-guided group

What is meta-learning model?



Double-blind, crossover RCT (NCT05032651)



Primary outcome and Secondary outcome

- **Primary outcome**: maintain **Hb near 11 g/dL within ± 0.25 g/dL**
- **Secondary outcome**: maintain **Hb between 10 - 12 g/dL within ± 15 %**

JAMIA Open, 2025, 8(2), ooaf020
https://doi.org/10.1093/jamiaopen/ooaf020
Research and Applications

JAMIA Open. 2025 Mar 27;8(2):ooaf020

Research and Applications

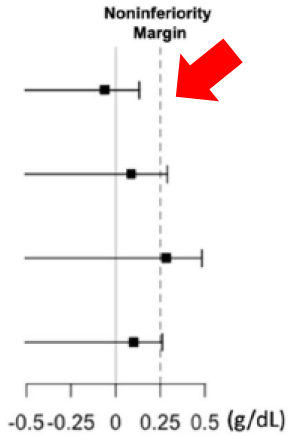
Computer-assisted prescription of erythropoiesis-stimulating agents in patients undergoing maintenance hemodialysis: a randomized control trial for artificial intelligence model selection

Lee-Moay Lim, MD[1,2], Ming-Yen Lin, PhD[1], Chan Hsu, MS[3], Chantung Ku, MS[3], Yi-Pei Chen, MS[1], Yihuang Kang, PhD[3,*], Yi-Wen Chiu, MD[1,2,4,*]

Equivalence testing: two one-sided test (TOST) within a **linear mixed-effects crossover model** including **treatment, period, and sequence as fixed effects** and **participant as a random effect**.

A

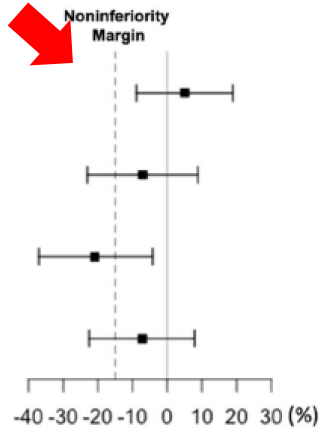
Model	Estimate	95% CI	p-value	Adjusted p-value
Bagging REEM Trees	-0.06	[-Inf, 0.13]	0.004	0.02
MERF	0.09	[-Inf, 0.29]	0.09	0.2
LSTM-I	0.28	[-Inf, 0.48]	0.6	0.6
LSTM-II	0.10	[-Inf, 0.26]	0.06	0.2



0.25 g/dL

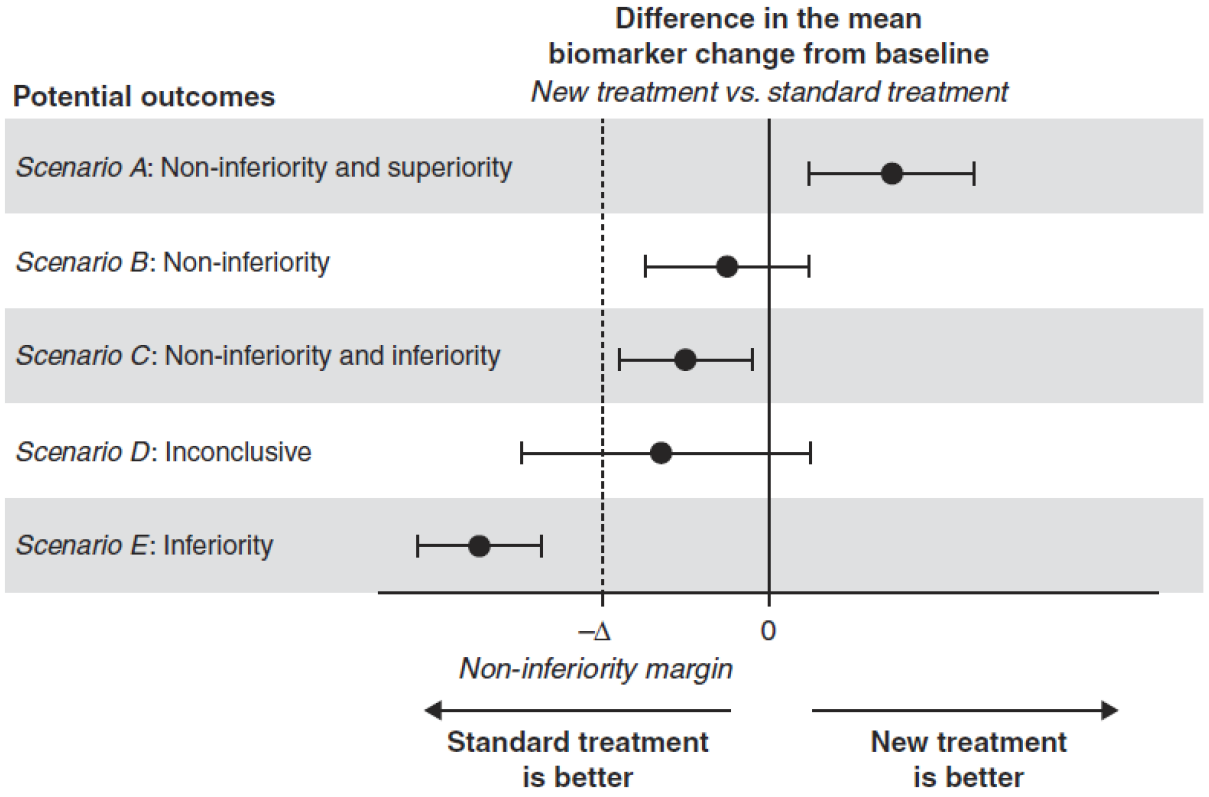
B

Model	Estimate	95% CI	p-value	Adjusted p-value
Bagging REEM Trees	5.00	[-8.88, 18.93]	0.003	0.01
MERF	-7.09	[-23.03, 8.82]	0.2	0.5
LSTM-I	-20.92	[-37.05, -4.19]	0.8	0.8
LSTM-II	-7.19	[-22.57, 7.88]	0.2	0.5

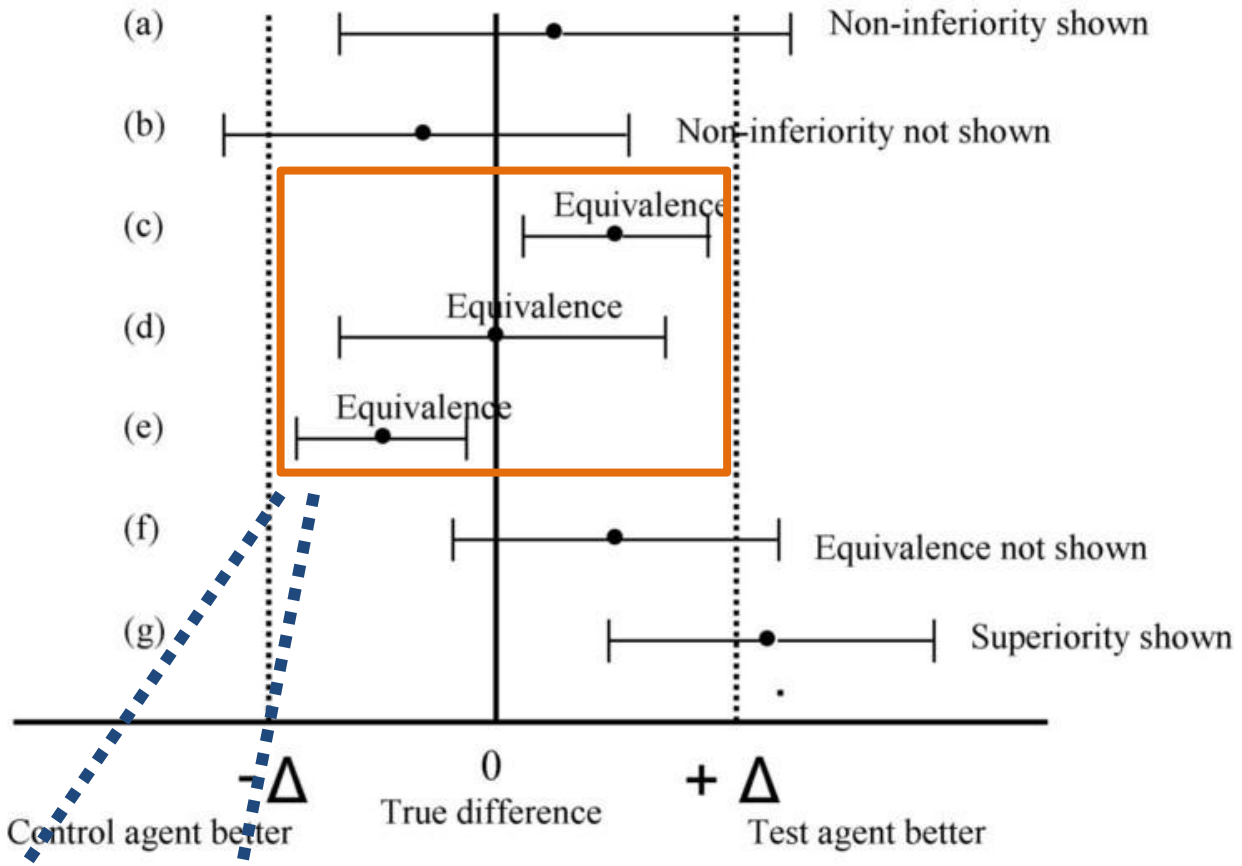


15 %

The concept of **Equivalence trial**



J Am Soc Nephrol. 2022 Apr;33(4):674-676



If the 95% CI of the mean difference in the achievement rate between "AI group – physician group" falls entirely **within –15% to +15%**, the two are considered **equivalent** in terms of "maintaining Hb within 10–12 g/dL."

If the 95% CI of the mean difference in this deviation value between "AI group – physician group" falls entirely **within –0.25 to +0.25 g/dL**, the two can be considered **"equivalent"** in terms of precision in Hb control.

Baseline Characteristics of Study Subjects

155 HD patients

- Mean age: **64.2** ± 12.5 years
- **Female** proportion: **45.8%**
- **Diabetes**: **44.5%**
- **Baseline Hb**: **10.8** ± 0.7 g/dL
- **Proportion with baseline Hb within 10–12 g/dL: 80.4%**
- **Completion Status at Each Phase:**
  - Arm 1 (enrolled 77 patients):
    - Phase I completers: 70 patients
    - Phase II completers: 63 patients
  - Arm 2 (enrolled 78 patients):
    - Phase I completers: 75 patients
    - Phase II completers: 71 patients
- **No significant difference** in **dropout rates** between groups (p = 0.11)
- **Serious adverse events (SAEs)** were **similar** between groups (p=0.63)

Baseline Participant Characteristics

	Arm 1 (N=77)	Arm 2 (N=78)	P-value
Age	66.2 (12.3)	62.2 (13)	0.054
Gender, female, %	40, 51.9%	31, 39.7%	0.148
DM, n, %	34, 44.2%	35, 44.9%	1
Hb, g/dL	**10.8 (0.7)**	**10.8 (0.7)**	0.572
Ferritin, ng/mL	452.9 (292.7)	504.3 (283.9)	0.269
Exclusion, n, %	14, 18.2	7, 9.0	0.218
Death	2, 2.6	0, 0	
Cancer	0, 0	2, 2.6	
GI Bleeding	2, 2.6	0, 0	
Transfer	2, 2.6	0, 0	
Transfusion	8, 10.4	5, 6.4	

Figure 1 Primary outcome across all assessments

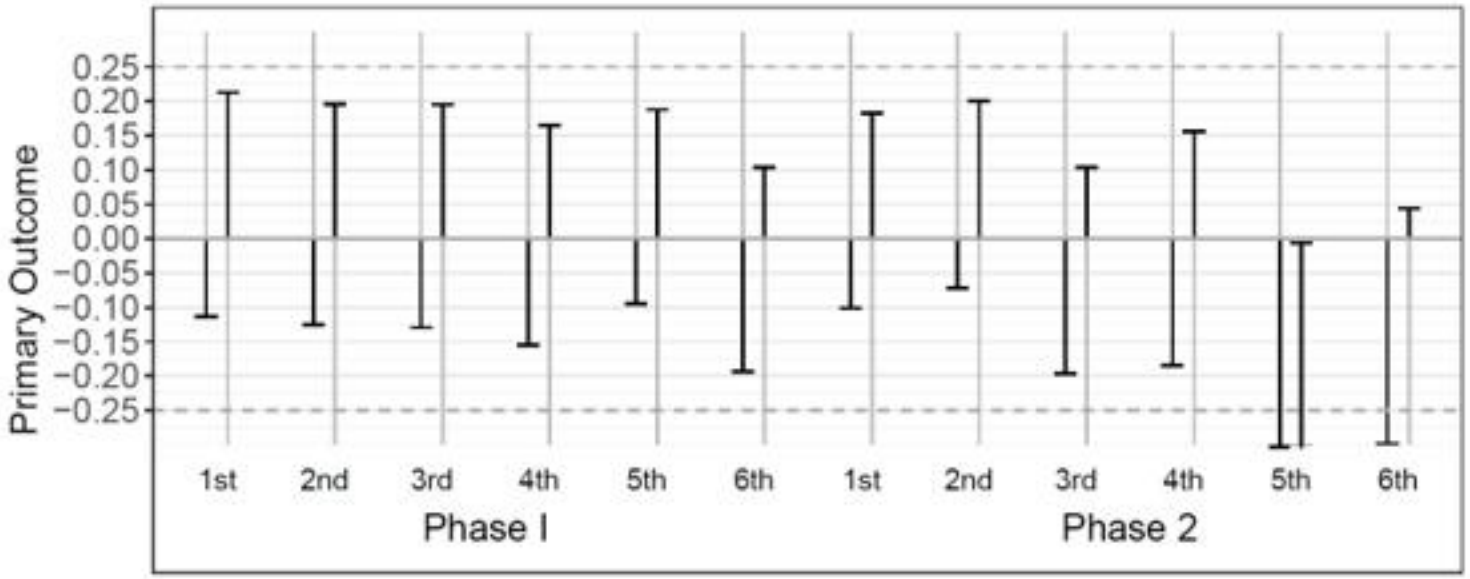


Figure 1. **Mean differences in absolute hemoglobin deviation from the 11 g/dL target (AI – Physician) across 12 assessments**. Error bars represent 95% confidence intervals from the linear mixed-effects model. The dashed lines denote the predefined equivalence bounds (±0.25 g/dL). At the 5th and 6th assessments in Phase 2, the lower bounds slightly exceeded –0.25 g/dL, suggesting transient AI superiority, but the overall model confirmed statistical equivalence between groups.

Figure 2 Secondary outcome across all assessments

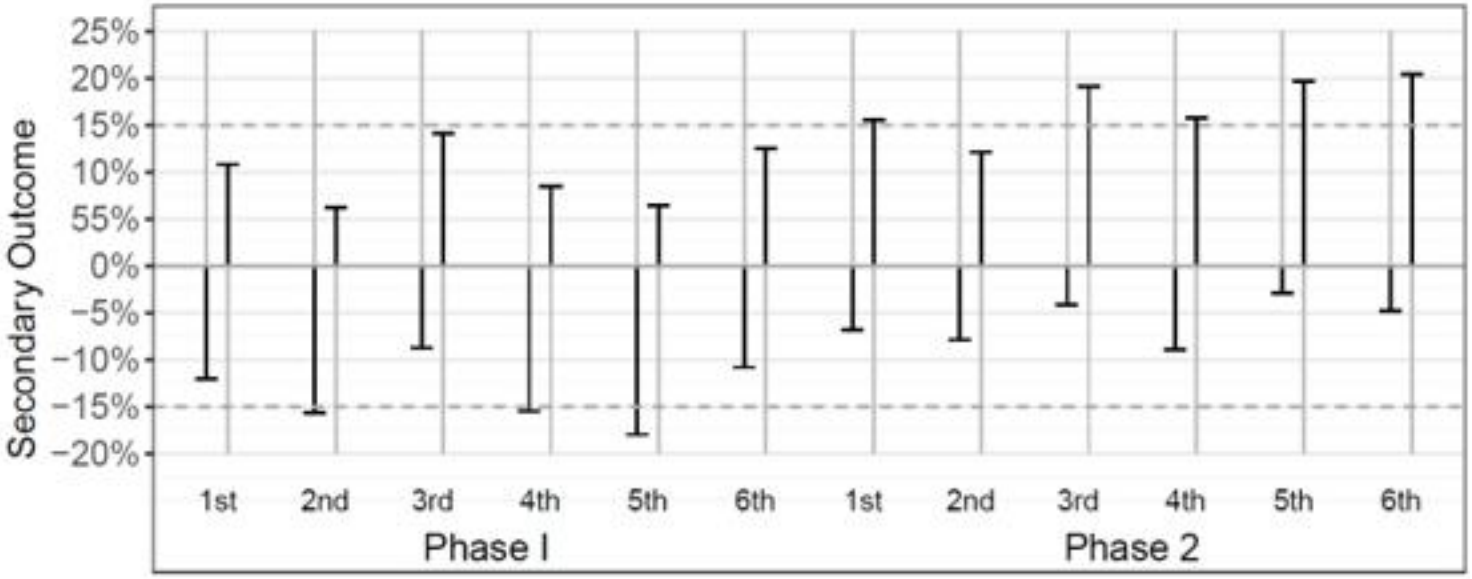
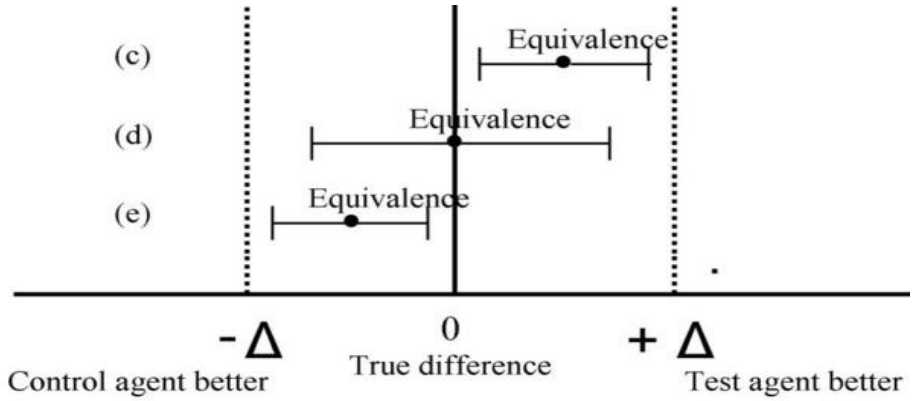


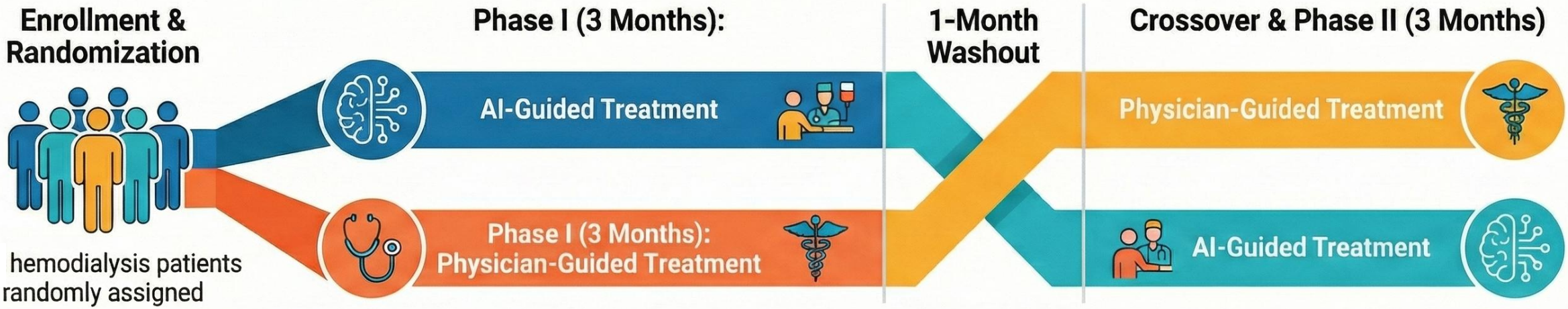
Figure 2. **Mean differences in maintaining Hb within 10–12 g/dL (AI – Physician) across 12 assessments**. Error bars show 95% CIs; dashed lines mark ±15% equivalence bounds. In Phase 2, 3rd–6th assessments exceeded the upper bound, suggesting transient AI overperformance, but overall analysis confirmed equivalence.

Primary and secondary outcomes

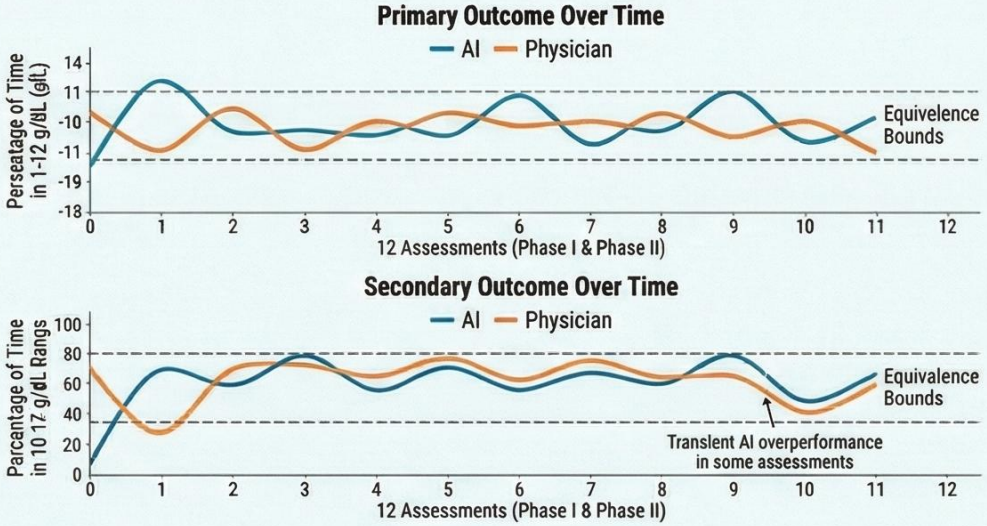
	AI Group	Dr Group	Two One-sided test (upper bound) 95% CI	Two One-sided test (lower bound) 95% CI
Overall	N=862 tests	N=853 tests		
Primary Outcome, mg/dL	**0.68±0.57**	**0.69±0.55**	~ to 0.03 (p<0.001)	-0.05 to ~(p<0.001)
Secondary Outcome, %	**77.7%**	**76.0%**	~ to 0.05 (p<0.001)	-0.01 to ~(p<0.001)
Phase I	N=436 tests	N=465 tests		
Primary Outcome, mg/dL	**0.72±0.60**	**0.70±0.56**	~ to 0.08 (p<0.001)	-0.04 to ~ (p<0.001)
Secondary Outcome, %	**74.1%**	**75.9%**	~ to 0.03 (p<0.001)	-0.07 to ~ (p<0.001)
Phase II	N=426 tests	N=388 tests		
Primary Outcome, mg/dL	**0.65±0.54**	**0.69±0.54**	~ to 0.02 (p<0.001)	-0.10 to ~ (p<0.001)
Secondary Outcome, %	**81.5%**	**76%**	~ to 0.10 (p<0.001)	-0.01 to ~ (p<0.001)

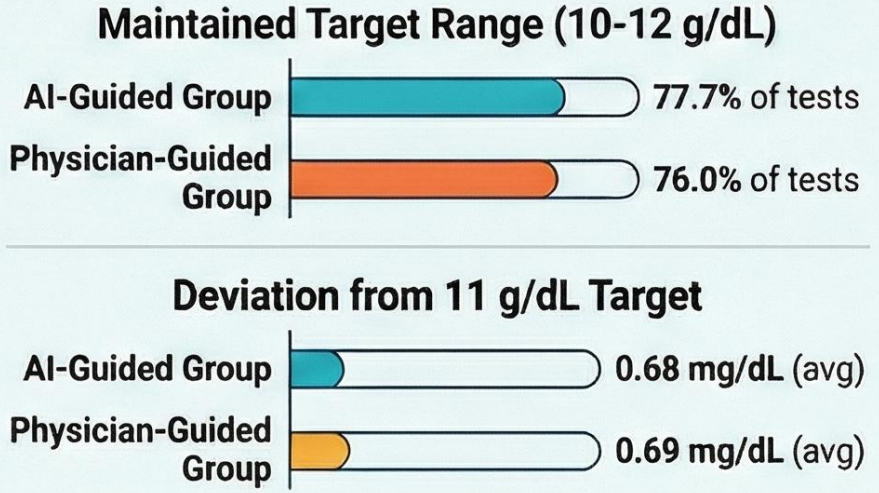


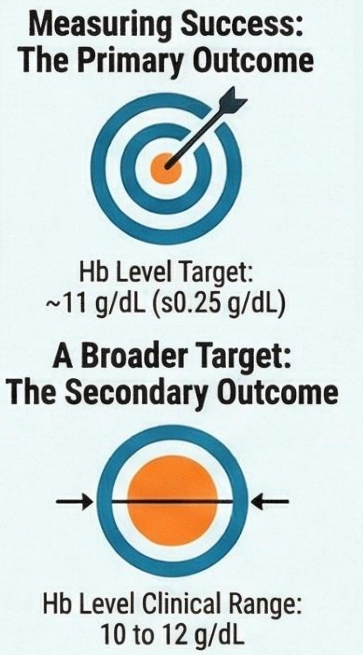
AI vs. Physician: A Tie in Anemia Management for Hemodialysis Patients



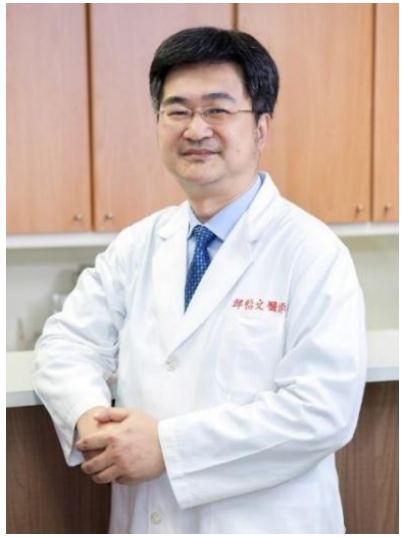
The Verdict: AI Achieves Equivalence







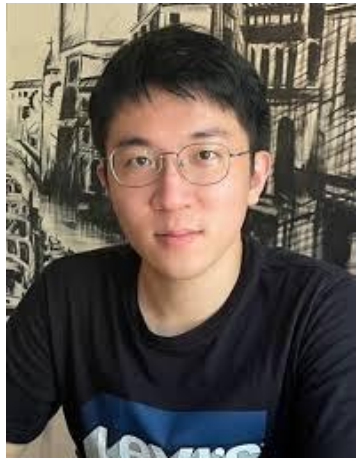
Acknowledgement



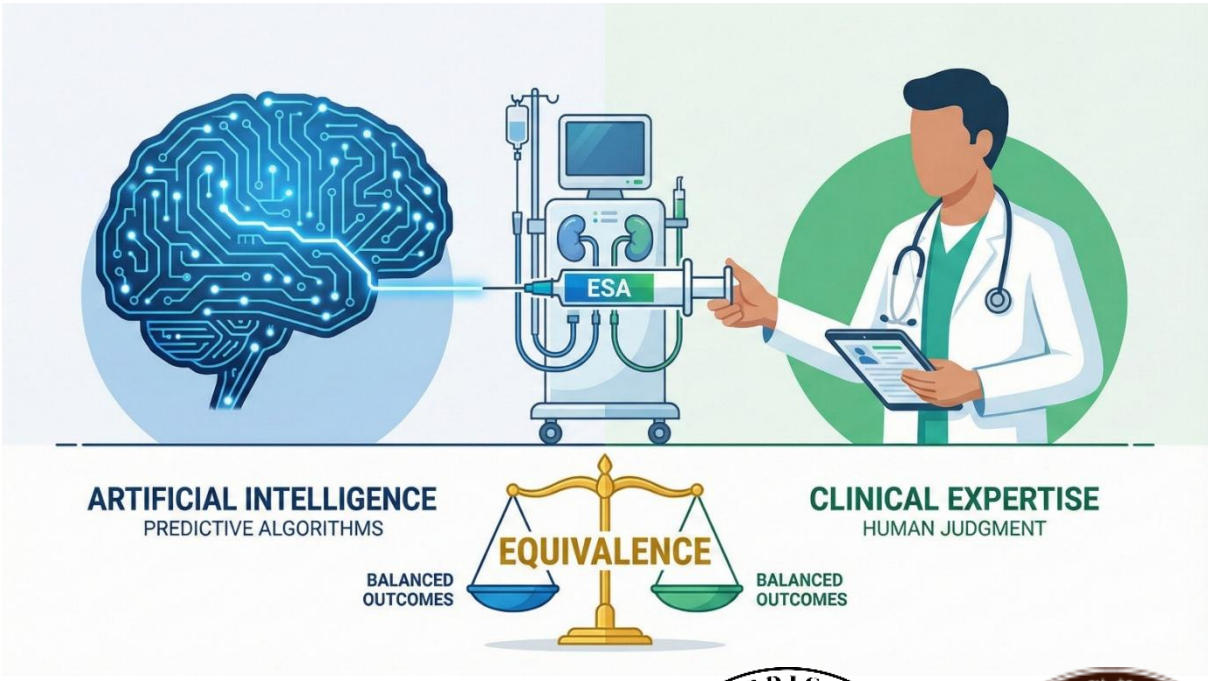
Yi-Wen Chiu, Prof
KMUH

Yihuang Kang, Assoc. Prof
NSYSU

Ming-Yen Lin, Assoc. Researcher
KMUH



Chan Hsu, PhD Student
NSYSU



Ping-Hsun Wu, M.D. PhD.
E-mail: 970392KMUH@gmail.com