Standardization sessions improved urologists' agreement and boosted AI consistency in uroflowmetry interpretation.

Plata M[1], Azuero G[1], Daza F[1], Gutierrez A[1], Garcia V[1], Rojas-Rivillas M[1], Gomez S[2], Rodriguez T[2], Cañar S[2], Avila P[2], Gonzalez S[2], Florez S[2], Giraldo L[2]

[1]Department of Urology, Fundación Santa Fe de Bogotá
[2]Universidad de los Andes

Improving Uroflowmetry Interpretation: Effects of Standardization Sessions on Interobserver Agreement and AI Model Consistency

Background
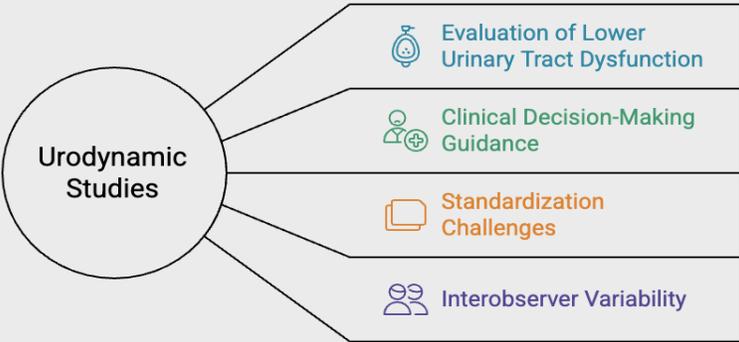
This study aimed to **assess** the **impact** of **standardization** sessions on **interobserver agreement** in **uroflowmetry** and their effect on **AI performance** and consistency.

Methods

- Weekly standardization sessions to review key urodynamic parameters until consensus.
- 50 uroflowmetries re-evaluated by 3 urologists pre- and post-standardization.
- Agreement: intraobserver (Fleiss' κ).
- AI: decision-tree models trained pre and post-standardization (F1-score).
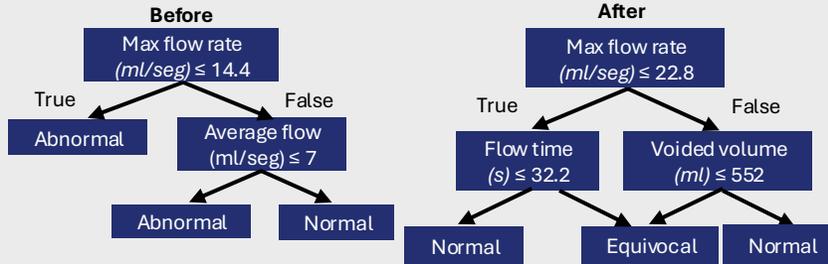- $p < 0.05$ significant.

Fleiss' kappa index

0 – 0.20	0.41 – 0.60	0.61 – 0.80
Slight	Moderate	Substantial

Results

Table 1. Interobserver agreement

	Pre-standarization	Post-standarization
	Kappa	
Uroflowmetry	0.5	0.772
Curve morphology	0.5	0.513

Figure 1. Impact of standardization on AI decision-tree models

AI model mean **F1-score:** increased from 0.62 to **0.67**

Implications

- Standarization **reduces** interobserver **variability.**
- Greater agreement among urologists enhances AI performance and consistency.
- This dual benefit underscores the **potential** of integrating **human expertise** with **AI** to advance the **reliability** and **standardization** of urodynamic study interpretation.