

# Revisiting the Cut-Off for an Abnormal Balloon Expulsion Test: Can We Make Testing More User Friendly?

Kailynn Barton<sup>1</sup>

William D. Chey MD<sup>1</sup>

Eric Shah MD<sup>1</sup>

Stacy Menees MD<sup>1</sup>

Jason R. Baker PhD<sup>1</sup>

<sup>1</sup> University of Michigan, Michigan Bowel Control Program



**MICHIGAN MEDICINE**  
UNIVERSITY OF MICHIGAN

# Background

- The prevalence of chronic constipation worldwide for adults is 16%<sup>1</sup>
  - Higher prevalence in females and older adults
- Anorectal Function Testing is used to identify constipated patients with an Evacuation Disorder including Dyssynergic Defecation (DD)
  - Of the available tests, Balloon Expulsion Testing (BET) is easiest, most accessible, and least costly
- Lack of consensus on the ideal Balloon Expulsion Test (BET) cutoff time which identifies a patient with an Evacuation Disorder
  - 60 seconds is the most commonly used cutoff time<sup>2</sup>
  - Others have argued that 2 minutes may improve accuracy<sup>3</sup>
- **Goal: Determine the optimal cutoff time for an abnormal BET using anorectal manometry and defecography as a gold standard**

1. Forootan M, Bagheri N, Darvishi M. Chronic constipation: A review of literature. *Medicine (Baltimore)*. 2018;97(20):e10631.

2. Shah ED, Farida JD, Menees S, Baker JR, Chey WD. Examining Balloon Expulsion Testing as an Office-Based, Screening Test for Dyssynergic Defecation: A Systematic Review and Meta-Analysis. *Am J Gastroenterol*. 2018;113(11):1613-1620.

3. Chiarioni G, Pieramico O, Vantini I, Heymen S, Whitehead WE. Utility of the Balloon-Evacuation Test for Identifying Patients With Dyssynergic Defecation. *Gastroenterology*. 2011;140:S797.

# Objective

- **Null Hypothesis**: The optimal threshold for a normal BET time is 60 seconds.
- **Alternative Hypothesis**: The optimal threshold for a normal BET time is not 60 seconds.

# Methods: Data Collection

- Retrospective evaluation of 4746 CC patients who underwent anorectal function testing at Michigan Medicine from 7/03 – 2/20
- Diagnostic test data evaluated:
  - BET time
  - Anorectal Manometry (ARM)
  - Defecography (DEF)
  - Dyssynergic Defecation (DD) defined as an abnormal ARM and DEF results (Rome IV)
- BET cut-offs were evaluated by 30 second increments

# Methods: Definitions of Anorectal Function Tests

- BET
  - Normal BET: The ability to expel a 50 mL water-filled balloon in  $\leq 60$  seconds
  - Abnormal BET: The inability to expel a 50 mL water-filled balloon in  $\leq 60$  seconds
- ARM
  - Normal ARM: Sphincter relaxation during simulated defecation
  - Abnormal ARM: Sphincter contraction or failure to relax during simulated defecation
- DEF
  - Normal DEF: During defecation: relaxation of the anal sphincter & puborectalis accompanied by evacuation of barium paste from the rectum
  - Abnormal DEF: During defecation: contraction or lack of relaxation of the sphincter or puborectalis and/or inability to expel barium paste from the rectum
- DD
  - Normal DD: Normal ARM and DEF results
  - Abnormal DD: Abnormal ARM and DEF results

# Methods: Statistical Measures

- **Definitions**

- **Sensitivity**: Probability of an abnormal BET time relative to ARM, DEF, and DD
- **Specificity**: Probability of a normal BET time relative to ARM, DEF, and DD
- **PPV**: Probability that the patient has an abnormal ARM, DEF, or DD when the BET time is abnormal
- **NPV**: Probability that the patient has a normal ARM, DEF, or DD when the BET time is normal

- **Statistical analysis:**

- **Binary Classification Test:** Sensitivity, Specificity, PPV, and NPV
  - BET results based on different time cutoffs (30-second increments) vs. ARM, DEF, or DD results

# Results: Study Cohort Demographics

<b>Variable</b>	<b>Value</b>
<b>Mean Age</b>	<b>48.8±16.4 years (18-93)</b>
<b>Mean BMI</b>	<b>27.5±7.2</b>
<b>Sex</b>	<b>Female = 81.8%</b>
<b>Race</b>	<b>Caucasian = 85.3%</b> <b>African American = 9.4%</b> <b>Other = 5.3%</b>

# Results: BET vs. Anorectal Manometry (ARM)

BET Time Points	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)
30-Seconds vs. Abnormal ARM	59.2% (57.0%, 61.4%)	51.3% (49.4%, 53.2%)	47.7% (46.3%, 49.0%)	62.7% (61.1%, 64.2%)
60-Seconds vs. Abnormal ARM	58.9% (57.1%, 60.9%)	54.2% (52.0%, 56.4%)	62.1% (60.8%, 63.5%)	50.9% (49.3%, 52.4%)
90-Seconds vs. Abnormal ARM	58.3% (56.4%, 60.1%)	55.1% (52.7%, 57.4%)	68.0% (66.7%, 69.3%)	44.6% (43.1%, 46.1%)
120 Seconds vs. Abnormal ARM	55.5% (53.9%, 57.2%)	54.3% (51.3%, 57.3%)	79.3% (78.2%, 80.5%)	27.9% (26.6%, 29.2%)



# Results: BET vs. Defecography (DEF)

BET Time Points	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)
30-Seconds vs. Abnormal DEF	63.8% (58.5%, 68.9%)	55.7% (50.4%, 60.9%)	57.7% (54.2%, 61.1%)	61.9% (57.9%, 65.8%)
60-Seconds vs. Abnormal DEF	61.9% (57.2%, 66.5%)	62.5% (56.6%, 68.2%)	71.8% (68.3%, 75.1%)	51.6% (47.9%, 55.3%)
90-Seconds vs. Abnormal DEF	60.4% (55.9%, 64.9%)	63.5% (57.1%, 69.6%)	76.3% (72.9%, 79.4%)	45.2% (41.6%, 48.8%)
120 Seconds vs. Abnormal DEF	53.4% (49.4%, 57.4%)	55.3% (44.7%, 65.6%)	88.8% (86.3%, 91.0%)	15.2% (12.8%, 17.9%)

# Results: BET vs. Dyssynergic Defecation (DD)

BET Time Points	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)
30-Seconds vs. Abnormal DD	83.2% (78.8%, 87.1%)	35.5% (30.7%, 40.6%)	53.8% (51.6%, 56.0%)	70.2% (64.2%, 70.6%)
60-Seconds vs. Abnormal DD	82.3% (78.4%, 85.8%)	40.4% (34.7%, 46.4%)	68.1% (65.7%, 70.3%)	59.7% (53.6%, 65.5%)
90-Seconds vs. Abnormal DD	80.8% (77.0%, 84.3%)	41.5% (35.1%, 48.1%)	72.8% (70.5%, 75.0%)	52.4% (46.4%, 58.2%)
120 Seconds vs. Abnormal DD	73.5% (69.9%, 76.9%)	27.7% (18.9%, 37.9%)	87.1% (85.5%, 88.5%)	13.6% (10.0%, 18.3%)

# Study Limitations

- **Retrospective**
- **Single, tertiary-care referral center**
- **Cross-sectional study design**
  - **Lack of patient follow-up data**
- **Known rate of abnormal simulated defecation responses on ARM in *healthy* persons (M>F)**
- **Did not control for medication(s) that could affect BET, ARM, or DEF results**

# Conclusions

- **A BET threshold of  $\leq 60$  seconds offers the optimal blend of performance characteristics for identifying patients with DD when using ARM and/or DEF as a gold standard.**
- **This argues for the continued use of a 60-second time cutoff when performing BET.**
- **The inability to pass a balloon in 90 or 120 seconds increases the diagnostic certainty for DD.**
- **GI physiology laboratories performing BET should consider using 60 seconds as the time cutoff for a normal test but collecting data for 120 seconds to improve diagnostic certainty for DD.**