

# Dataset Integrity Check for Evaluating Predictors and Interventions in Sphincter of Oddi Dysfunction (EPISOD)

**Prepared by Adam Risch**

**IMS Inc.**

3901 Calverton Blvd, Suite 200 Calverton, MD 20705

**April 9, 2019**

## Contents

1 Standard Disclaimer .....	2
2 Study Background .....	2
4 Statistical Methods .....	2
5 Results .....	3
For Table 1 in the publication [1], The EPISOD study: long-term outcomes, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data files to the results published in Table 1. The results of the replication are identical to the published results. ....	3
6 Conclusions .....	3
7 References .....	3
Table A: Variables used to replicate Table 1: Success rates of individuals originally enrolled in EPISOD and in the long-term follow-up study .....	4
Table B: Comparison of values computed in integrity check to reference article Table 1 values .....	5
Attachment A: SAS Code .....	6

## **1 Standard Disclaimer**

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is not to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

## **2 Study Background**

The EPISOD Trial is a two-arm parallel, randomized, doubleblinded, sham-controlled, multicenter Phase III clinical trial of endoscopic sphincterotomy as treatment for adults 18 to 65 years of age diagnosed with SOD III.

## **3 Archived Datasets**

All the SAS data files, as provided by the Data Coordinating Center (DCC), are located in the EPISOD folder in the data package. For this replication, variables were taken from the “derived.sas7bdat” datasets in the ““/EPISOD/private\_orig\_data/2\_2\_2018” and “/EPISOD/private\_orig\_data/PUDS/SAS files” folders.

## **4 Statistical Methods**

Analyses were performed to duplicate results for the data published by Cotton et al [1] in *Gastrointestinal Endoscopy* Jan 2018, Volume 87, Issue 1. To verify the integrity of the dataset, descriptive statistics were computed.

## 5 Results

For Table 1 in the publication [1], The EPISOD study: long-term outcomes, Table A lists the variables that were used in the replication and Table B compares the results calculated from the archived data files to the results published in Table 1. The results of the replication are identical to the published results.

## 6 Conclusions

The NIDDK repository is confident that the EPISOD 2\_2\_2018 data files to be distributed are a true copy of the study data.

## 7 References

[1] Peter B. Cotton, MD, FRCS, FRCP, Qi Pauls, MS, Jamila Keith, BS, Andre Thornhill, MS, Douglas Drossman, MD, April Williams, BS, Valerie Durkalski-Mauldin, MPH PhD. "The EPISOD Study: long term outcomes". *Gastrointestinal Endoscopy*, Volume 87 , Issue 1 , 205 – 210.

**Table A:** Variables used to replicate Table 1: Success rates of individuals originally enrolled in EPISOD and in the long-term follow-up study

<b>Table Variable</b>	<b>dataset.variable</b>
Sphincterectomy arm	derived.ztreatmentnm
Sham arm	derived.ztreatmentnm
EPISOD cohort	number of observations in PUDS/SAS files/derived.sas7bdat
Success at 1 year by RAPID (EPISOD Cohort)	PUDS/SAS files/derived.f2, f3, & f4
Long-term enrollees	number of observations in 2_2_2018/derived.sas7bdat
Success at 1 year only by RAPID (Long-term enrollees)	derived.status
Success at 1 year and final visit by RAPID (Long-term enrollees)	derived.epi15_s
Success at final visit only by RAPID (Long-term enrollees)	derived.epi5_lv_s
Individuals with PGIC data	derived.epi15_pgic_s & epi5_lv_pgic_s
Success by PGIC at 1 year and final visit	derived.epi15_pgic_s
Success by PGIC at final visit only	derived.epi5_lv_pgic_s
Success by RAPID at 1 year and final visit (subjects w/ PGIC data)	derived.epi15_s
Success by RAPID at final visit only (subjects w/ PGIC data)	derived.epi5_lv_s

**Table B:** Comparison of values computed in integrity check to reference article Table 1 values

	Manuscript (N = 7914)			DSIC (N = 7914)			Diff. (N = 0)		
	Sphincterectomy arm, n (%)	Sham arm, n (%)	Unadjusted 95% CI for difference	Sphincterectomy arm, n (%)	Sham arm, n (%)	Unadjusted 95% CI for difference	Sphincterectomy arm, n (%)	Sham arm, n (%)	Unadjusted 95% CI for difference
<b>EPISOD cohort</b>	11	73		141	73		0	0	
<b>Success at 1 year by RAPID</b>	32 (23)	27 (37)	1.2% to 27.3%	32 (23)	27 (37)	1.2% to 27.3%	0	0	0
<b>Long-term enrollees</b>	65	38		65	38		0	0	
<b>Success at 1 year only</b>	12 (18)	18 (47)	10.4% to 47.4%	12 (18)	18 (47)	10.4% to 47.4%	0	0	0
<b>At 1 year and final visit</b>	9 (14)	11 (29)	-1.6% to 31.8%	9 (14)	11 (29)	-1.6% to 31.8%	0	0	0
<b>At final visit only</b>	26 (40)	16 (42)	-17.6% to 21.8%	26 (40)	16 (42)	-17.6% to 21.8%	0	0	0
<b>Individuals with PGIC data</b>	43	22		43	22		0	0	
<b>Success by PGIC</b>									
<b>At 1 year and final visit</b>	6 (14)	11 (50)	12.7% to 59.4%	6 (14)	11 (50)	12.7% to 59.4%	0	0	0
<b>At final visit only</b>	16 (37)	16 (73)	12.0% to 59.1%	16 (37)	16 (73)	12.0% to 59.1%	0	0	0
<b>Success by RAPID</b>									
<b>At 1 year and final visit</b>	6 (14)	9 (41)	3.9% to 50.0%	6 (14)	9 (41)	4.0% to 50.0%	0	0	0.1
<b>At final visit only</b>	18 (42)	13 (59)	-8.1% to 42.5%	18 (42)	13 (59)	-8.1% to 42.5%	0	0	0

## Attachment A: SAS Code

```
/* Program Documentation
/** Program Name: /prj/niddk/ims_analysis/EPISOD/prog_initial_analysis/dsic_2_2_2018.sas
/** Author: Risch
/** Date: 3/7/2019
/** Brief Description:
Performs a data set integrity check for the EPISOD "2_2_2018" data submission and table 1 of the associated primary paper.
/** Additional Description:
/** End Description:
*/

* Input files;;
libname in '/prj/niddk/ims_analysis/EPISOD/private_orig_data/2_2_2018/';
libname puds '/prj/niddk/ims_analysis/EPISOD/private_orig_data/PUDS/SAS files/';
* End Input;

* Output files;;

* End Output;

title 'DSIC for EPISOD 2-2-2018 Data Submission (/prj/niddk/ims_analysis/EPISOD/prog_initial_analysis/dsic_2_2_2018.sas)';

options nocenter validvarname=uppercase missing=' ' nofmterr linesize=256 /*mprint*/;

proc format;
  value $ arm
    'Biliary','Dual' = 'Sphincterectomy'
    'Sham' = 'Sham';
;

data codelist;
  set in.codelist;

data datadictionary;
  set in.datadictionary;

data derived;
  set in.derived;

data form12;
  set in.form12;
```

```

data form14;
  set in.form14;

data form15;
  set in.form15;

data form16;
  set in.form16;

data form50;
  set in.form50;

data form51;
  set in.form51;

data form52;
  set in.form52;

data form53;
  set in.form53;

data form54;
  set in.form54;

data form55;
  set in.form55;

data puds_derived;
  set puds.derived;
  if not (f2 | f3 | f4) then rapid_lyr_success = 1;
  else rapid_lyr_success = 0;

proc contents data=_all_;

proc freq data=puds_derived(where=(ztreatmentnm ne ''));
  title2 'EPISOD Cohort';
  tables ztreatmentnm / missing;
  tables ztreatmentnm * rapid_lyr_success / cl missing;
  format ztreatmentnm $arm.;

proc freq data=derived;
  title2 'Long Term Enrollees';
  tables ztreatmentnm / missing;
  format ztreatmentnm $arm.;

proc freq data=derived(where=(ztreatmentnm ne ''));
  title2 'Long Term Enrollees';
  tables ztreatmentnm * (status epi15_s epi5_lv_s) / cl missing;
  format ztreatmentnm $arm.;

proc freq data=derived(where=(ztreatmentnm ne '' and (epi15_pgic_s ne . or epi5_lv_pgic_s ne .)));

```



```
title2 'Long Term Enrollees w/ PGIC data';  
tables ztreatmentnm / missing;  
tables ztreatmentnm * (epi15_pgic_s epi5_lv_pgic_s epi15_s epi5_lv_s) / cl missing;  
format ztreatmentnm $arm.;
```