



CORI v4 National Endoscopic Database

The data in the National Endoscopic Database (NED) is captured as part of the documentation of endoscopic procedures. This document describes the structure and data in the CORI v4 NED, i.e. data captured by documentation using the CORI v4 Endoscopic Reporting Software. Please see associated documents for a broader view of this database (Key Points about the National Endoscopic Database, and Overview of the National Endoscopic Database) as well as documentation of the CORI v3 National Endoscopic Database.

The main procedure data table is the `proc_main` table.

- The Proc_Main table contains a globally unique procedure identifier (*procedure_uid*) for every procedure in the NED. This identifier links procedure data in the Proc_Main table to all other data in the NED for a given procedure; in all other data tables, the *procedure_uid* field functions as a foreign key to the Proc_Main table.
- A second procedure identifier, *proc_num*, is available and was assigned because it is easier to use for research purposes than *procedure_uid*. However, this identifier is found only in the Proc_Main table.
- The Proc_Main table also contains a unique patient identifier (*patient_uid*) associated with each procedure and a site identifier (*site_id*) which link to the Patient and Sites tables, respectively.
- The Proc_Main table also identifies the type of procedure that was documented (*procedure_type*). The procedure type is often important for linking the Proc_Main procedure data to the correct data tables when looking for detailed procedure data. Many procedure data tables are named with a three, four, or five letter prefix indicating the procedure type.
- Each of these prefixed procedure data tables has a *procedure_uid* field that contains the unique procedure identifier. Some tables may contain more than one row for a procedure. For instance, if multiple polyps are found during a Colonoscopy, there will be multiple rows in the ColPolyp table with that procedure's *procedure_uid*.

- In such cases there is a *finding_uid* field. This field is a unique identifier for the specific finding (in the above example, for a specific polyp). The *procedure_uid* and the *finding_uid* taken together can be used to construct a composite key for that table.

Procedure types and naming conventions

The procedure types in the *procedure_type* column of *proc_main* are:

BRON – Bronchoscopy
 BRONp – Pediatric bronchoscopy
 CAP – Capsule endoscopy
 CAPp – Pediatric capsule endoscopy
 COL – Colonoscopy
 COLp – Pediatric colonoscopy
 EGD – Esophagogastroduodenoscopy (EGD)
 EGDp – Pediatric esophagogastroduodenoscopy (EGD)
 ERCP – Endoscopic retrograde cholangiopancreatography (ERCP)
 ERCPp – Pediatric endoscopic retrograde cholangiopancreatography (ERCP)
 EUS – Endoscopic ultrasound (EUS)
 EUSp – Pediatric endoscopic ultrasound (EUS)
 FLX – Flexible sigmoidoscopy
 FLXp – Pediatric flexible sigmoidoscopy
 MOT – Motility study
 MOTp – Pediatric motility study
 NON – Non-endoscopic procedure
 SCE – Small caliber endoscopy

These abbreviations prefix many table names. As an example of the naming conventions used in the CORI v4 NED, table names beginning with "Col" contain Colonoscopy data.

ColInd is a table containing the Indication data for Colonoscopies.

ColFind is a table containing the Findings data for Colonoscopies.

Similarly,

EgdInd is a table containing the Indication data for EGDs.

EgdFind is a table containing the Findings data for EGDs.

Pediatric procedure tables are usually prefixed with "PEDS", for example:

PEDS_ColInd is a table containing the Indication data for pediatric colonoscopies.

PEDS_COLFind is a table containing the Findings data for pediatric colonoscopies.

(Exception: The Bronchoscopy module does not have separate tables for pediatric and adult Bronchoscopy procedure types.)

Some table names do not have a procedure type prefix. These tables usually contain data from multiple procedure types. For instance, most therapeutic procedures (except those specific to ERCP, EUS and Bronchoscopy procedures) are documented in tables prefixed with "Tx" regardless of procedure type:

TxHemAPC is a table containing Hemostatic Argon Plasma Coagulation data for most procedure types, including Colonoscopies, EGDs, and others, including some Findings documented in the ERCP and EUS procedure types.

TxNonHemAPC is a table containing Non-hemostatic Argon Plasma Coagulation data for those same procedure types.

CORI v4 Screenshots and Table names

Screenshots from the CORI v4 Endoscopic Reporting Software have been provided by procedure type in the following documents:

<u>Document</u>	<u>Procedure type</u>
CORI4_COL_Screenshots.pdf	Colonoscopy
CORI4_COL_Peds_Screenshots.pdf	Pediatric Colonoscopy
CORI4_FLX_Screenshots.pdf	Flexible Sigmoidoscopy
CORI4_FLX_Peds_Screenshots.pdf	Pediatric Flexible Sigmoidoscopy
CORI4_EGD_Screenshots.pdf	EGD
CORI4_EGD_Peds_Screenshots.pdf	Pediatric EGD
CORI4_EUS_Screenshots.pdf	EUS
CORI4_EUS_Peds_Screenshots.pdf	Pediatric EUS
CORI4_ERCP_Screenshots.pdf	ERCP
CORI4_ERCP_Peds_Screenshots.pdf	Pediatric ERCP
CORI4_CAP_Screenshots.pdf	Capsule Endoscopy
CORI4_CAP_Peds_Screenshots.pdf	Pediatric Capsule Endoscopy
CORI4_MOT_Screenshots.pdf	Motility Study
CORI4_MOT_Peds_Screenshots.pdf	Pediatric Motility Study
CORI4_SCE_Screenshots.pdf	Small Caliber Endoscopy
CORI4_NON_Screenshots.pdf	Non-Endoscopic Procedure
CORI4_BRON_Screenshots.pdf	Bronchoscopy
CORI4_BRON_Peds_Screenshots.pdf	Pediatric Bronchoscopy
CORI4_Postproc_Screenshots.pdf	Postprocedure module used with all procedures

The main screens for each procedure type are the tabs along the left side of any procedure screen. For example, the main screens for Colonoscopy and tables for data found on those screens are:

<u>Screenshot Tab</u>	<u>Data table</u>
History	History
Physical Exam	PE
Liver Disease	Liver
Indications	ColInd
Preprocedure	ColPreProc
Sedation	ColSedation

<u>Screenshot Tab</u>	<u>Data table</u>
Procedure	ColProc
COL Findings	ColFind
Sm. Bowel Findings	ColSBFind
Events	Intervention
Assessment/Plan	TreatmentPlan
Letters/Instructions	Letters

Some screens, and therefore their associated data tables, are used for more than one procedure type.

Example: The Liver Disease screens are identical for all adult procedure types. Data for all adult procedure types which document liver disease will be found in a single Liver table. Similarly, data for all pediatric procedure types which document liver disease will be found in the PEDS_Liver table.

In addition to these main procedure tables, there are at least four other types of tables derived from procedure data:

- The data from each grid is stored in its own table.
Example: The Colonoscopy Preprocedure screen shows a grid for Procedure Personnel. That grid data will be found in the ColPreProc_Procedure_personnel_grid table. The screenshot documents give the name of each main table as well as the name of the grid tables.
- Details about Findings, Therapies, Indications, etc. are found in separate tables. The red arrows next to CheckFlag controls on the main screen indicate that additional details will be collected in a subscreen. Selecting a finding in the Finding menu also launches a subscreen. There is one table for each subscreen. Subscreens may be used for more than one procedure type.

Examples:

When the colonoscopy finding of Diverticulosis is selected, the Diverticulosis detail subscreen appears. Data from this window is in the ColDivertic table.

On the Diverticulosis detail subscreen, if the Hemostatic checkbox is checked, the Hemostatic Therapies menu subscreen appears. Data from this table is in the TheralauncherHemo table.

On the Hemostatic Therapies menu subscreen, if the APC checkbox is checked, the Hemostatic Argon Plasma Coagulation (APC) detail subscreen appears. Data from this window is in the TxHemAPC table.

- Pathology data is documented on the Postprocedure screen, but is always associated with a specific Finding. Any results documented will be found in a table related to that finding.

Examples:

The ColPolyp_Pathology table contains pathology data for Colonoscopy polyp findings.

The EGDUlcer_Pathology table contains pathology data for EGD ulcer findings.

- The Findings table lists each Finding that is documented for a procedure for all procedure types.

Missing tables

If no data has been entered into a screen, there will not be a table in the NED for that screen.

Data deletions

Text data has been redacted to remove protected health information, including patient or doctor names, full dates, and other personal identifiers. In these cases, the deleted string has been replaced with an indication that data has been deleted, such as [deleted], [date deleted], [redacted comment], or [name deleted]. Occasionally free-text fields (e.g. *comments* fields) have been removed entirely.

Pediatric data

Pediatric data (i.e. data on patients <18 years of age) may be found in non-pediatric tables as well. The location of the data depends on the procedure type selected by the staff at the procedure site. If staff members used the adult module to document a pediatric colonoscopy, the data for that colonoscopy will be in the adult ("Col") tables rather than the pediatric ("PEDS_Col") tables.

Endoscopist/Bronchoscopist demographic data

Endoscopist (and bronchoscopist) demographic data is stored in the CUEI table where the field *CUEI* (CORI4 Unique Endoscopist ID) contains a unique ID for each endoscopist. The CUEI links endoscopist demographic data to several other tables and fields. For example:

- The ColPreProc table contains data from the Preprocedure screen for Colonoscopy procedure types. The field *endoscopist* contains the CUEI for the responsible endoscopist and the *performer* field contains the CUEI for the endoscopist who performed the procedure.
- The Proc_Main table data for this procedure also contains a CUEI in the *CUEI* field for the endoscopist who signs the procedure, and who is usually, but not always, the same as the endoscopist who performed the procedure.

- The ColPreProc_Procedure_personnel_grid table associates the endoscopist's CUEI number (field *Name_c2*) with a specific role (e.g. "Endoscopist – Attending Physician", "Endoscopist – Fellow", "Endoscopist – Nurse Practitioner", etc.) in the procedure (field *Role_c1*).

Some roles and endoscopist/bronchoscopist fields may not have a value for the CUEI. This occurs in procedures which were not performed by endoscopists/bronchoscopists, such as many Non-endoscopic procedures.

The endoscopist may be linked to endoscopists in the CORIv3 NED through the *providerID* field in the CUEI table. If the endoscopist has data in the CORI v3 NED, the value in this field will be a 4-digit number which will link to the CORI3 dataset (Physician table, *physician* field). If the ProviderID value is a 9 digit number that matches the CUEI, this endoscopist does not have any associated procedures in the CORI v3 NED.

In 2015, more accurate certification data was sought for all physicians. Specialty boards were queried for current certifications. In some cases, such as when a physician was certified for a specialty outside of the USA, their certifications could not be determined. In these cases the *flag* field of the CUEI table will contain the value "not resolved". Previously, specialty was self-reported or reported by administrative staff.

Site data

Data about the endoscopy sites which have contributed data to the CORI v4 NED are included in the Sites table.

Patient demographics

Data about the patients is captured in the Patient table. There is one row in this table for each procedure and therefore, there may be more than one row in this table for each patient. The data in each row reflects the information known about the patient at the time of the procedure, some of which may change over time.

Patients can be identified within a site (or within a group where patients are shared) by a unique identifier, the *patient_uid*. Otherwise, patients cannot be identified across sites. If a patient receives endoscopic procedures at more than one site (not in the same group), that patient will have more than one identifier which cannot be linked.

Procedure and Diagnosis codes (CPT, ICD-9-CM)

Procedure (CPT) codes are assigned to the procedure or procedures performed when they are documented on the Procedure main screen for each procedure type. This assignment

is made based on a table in the CORI v4 application which can be customized by each site. The most recently suggested codes are found in the Cpt_Codes table. The assigned CPT codes are also found in the NED in fields which begin with the prefix "lbl", with one field for each procedure.

Examples:

If a Colonoscopy is performed, and the endoscopist selects "Colonoscopy" on the Colonoscopy Procedure screen, then in the ColProc table the field *colCPTColonoscopy* will have a value of either 1 or 2 (depending on documentation of more detailed procedures). In addition, the *lblColMainColonoscopy* field will contain the CPT code which the site has associated with this procedure.

If, in addition, a biopsy is performed and the endoscopist selects this option on the subscreen titled "Colonoscopy with ..." then in the ColProcColonoscopy table, the field *cptColBiopsy* will contain the value 1 and the field *lblCPTColBiopsy* will contain the assigned CPT code.

Diagnosis (ICD-9-CM) codes are associated with Findings. While ICD-10-CM codes were available to sites in late 2014, none of the sites had begun using them before the end of data collection for this dataset. When documenting a Finding detail, the endoscopist selects the most appropriate text for the endoscopic diagnosis for this finding, and may add free text. The ICD-9-CM code is then assigned based on a table in the CORI v4 application which can be customized by the site. The text selected or entered will be found in the field *diagnosis*, and the assigned code can be found in the *icd9* field for each Finding. The most recently suggested ICD-9-CM codes can be found in the table ICD9_codes.

AVAIL fields

In addition to the data fields in the NED, for every field with data captured using a control in the CORI v4 Endoscopic Reporting Software, there is a second field of the same name followed by "_avail". The _avail field tells the investigator whether or not the associated control was available and, if available, whether or not it was a required control in the application at the time the procedure was created. Required fields are a site-specific application setting, so a value of 0 or 1 may appear for the same application version, depending on which site is documenting a procedure.

Value	Meaning
1	Control available in the application version with which the procedure was documented
0	Control available and required in the application version in which the procedure was documented.

-1	Control NOT available in the application version in which the procedure was documented
null	No data in the parent field

Example: Through version 4.2.1.1, the level of supervision of an attending physician (when a fellow is present) in a colonoscopy procedure was noted with a checkflag control labelled "The attending was present during the procedure." Data from this control was captured in the ColPreProc table, *attendingPresentCol* field. After that version, the control was changed to a CComboBox labelled "Level of supervision" and with 3 options including "Attending was present for the entire procedure." Data from this control was captured in the ColPreProc table, *supervision* field.

- For a procedure documented using version 4.2.1.1, if there is data present in the *attendingPresentCol* field, then the *attendingPresentCol_avail* field will have a value of 1 (indicating that the *attendingPresentCol* control was available). If there is data in the *supervision* field, this data was added at a later date, so the *supervision_avail* field will have a value of -1 (indicating that the *supervision* control was not available for documentation at the time of the procedure, but data was later moved or entered into this field during a revision.)
- For a procedure documented using a version 4.2.1.2 or later, data can no longer be entered into the *attendingPresentCol*, so the *attendingPresentCol_avail* field will have a value of NULL. The associated *attendingPresentCol_avail* field will also be null since the *_avail* fields are only populated when the associated data column has data. The *supervision_avail* field will have a value of 1 (indicating that the *supervision* control was now available) or 0 (indicating that the supervision control was now available and required).

The data dictionary

A data dictionary is provided as an Excel spreadsheet. For each column in the CORI v4 NED, there is one row in the spreadsheet. The fields are described in the Table below.

Table: The CORI4 NED data dictionary

Column	Description
Study	CORI4
DataSet	The Table name
Variable	The field/column name
Type	SAS datatype
Length	SAS field length
SQL datatype	SQL database datatype
Label	The "pathway" to the control label on the screen, i.e. what the user saw when documenting on the screen

ControlType	The type of screen control for capturing the data - textbox, numeric field, option group, checkbox, menu, editable menu, etc. Note most of these are specialized control types for the CORI application.
Format	The possible values for the variable. Consider the following, however: <ul style="list-style-type: none"> • A few of the menus used in CORI v4 may be editable menus i.e while a list of suggested values is presented to the user the user may enter any value into the field. There are significantly less of these than in CORI v3, however. • It is not possible to know what the suggested value list was at the time of the procedure, so values listed in this field are probable but not absolute. • Examination of the values found in the field will be necessary in order to determine the possible values.
Screenshot	The name of the screenshot document illustrating the CORI v4 screens. When screens are used for more than one procedure type this will be noted.
Page	The page in the screenshot document illustrating the variable when available.
Notes	Notes about the variable that may help during data analysis.