

Overview of the National Endoscopic Database

The National Endoscopic Database (NED) contains data collected by the Clinical Outcomes Research Initiative (CORI) over a 15-year period, from 2000 to 2014. CORI is a research group associated with Oregon Health & Science University (OHSU) which was established in order to study outcomes of gastrointestinal (GI) endoscopic procedures in "real life" settings. To do this, an electronic health record was developed which allowed practitioners to document their procedures electronically. This is the CORI Endoscopic Reporting Software, which has two major versions, v3 and v4.

The CORI Endoscopic Reporting Software was licensed to participating endoscopic sites for use in routine clinical care. The sites in turn agreed to document a minimum of 95% of all endoscopic procedures using the software. Data collected in this manner was then returned daily to the CORI Data Center for inclusion in the NED. The two versions of the software have resulted in two databases which together make up the NED, one for each version. Although these two databases include data of essentially the same breadth and depth, their data structures differ significantly.

1. Participating Sites

Figure 1 shows a map of the sites that were participating with CORI in 2005. Sites have joined and left the consortium over the years and Figure 2 shows the count of sites by year.

Sites are identified by a unique ID and cannot be reidentified from the NED. Other data available about each site includes the state and region of the USA where the site is located, the site type (university, community, VA, military, HMO), the facility type (ambulatory center, hospital, office) and the dates of participation with CORI.

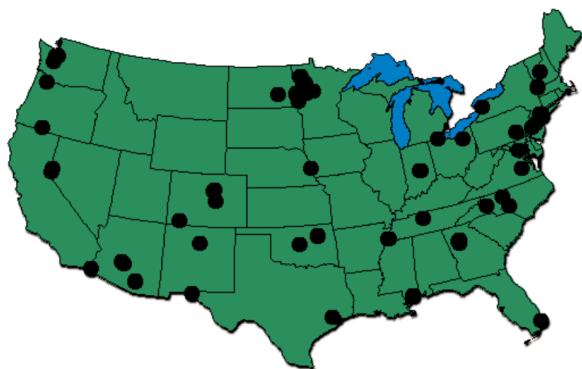


Figure 1. Site map, 2005

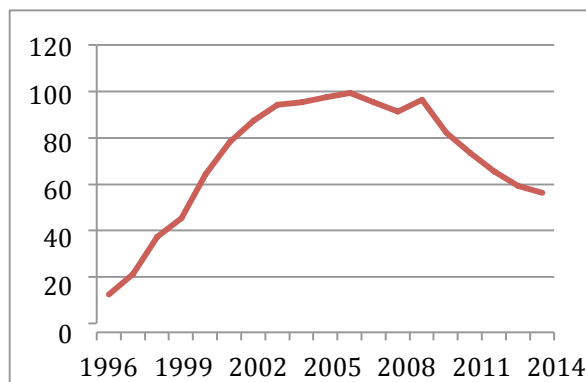


Figure 2. Site count by procedure year

Grouping of sites. The provider practices which participate in the CORI consortium have a variety of administrative structures. Some practices perform endoscopies at only one physical site. However, the CORI software server can be configured for use at more than one physical site; when this is the case, the sites are grouped with a common group identifier. These grouped sites may or may not share patients and/or providers.

For example, an academic medical center could have a single server, and therefore a single group ID, but also have 3 sites: a pediatric endoscopy unit, an outpatient adult endoscopy unit and an inpatient adult endoscopy unit. Each of these sites would have a unique site ID. The adult units might share providers and patients with each other but not with the pediatric unit.

Caveats on the use of site data.

- The sites included in the NED are not necessarily representative of all GI practices in US. Veterans Affairs facilities and Academic Medical Centers are over-represented, for example.
- Studies comparing sites have not been performed, partly due to the complexity of administrative structures of practices, groups and sites.
- Not all procedures at a site have to be entered into the CORI software. Some sites require all endoscopists at their facility to document in the CORI software and others do not.
- In the v3 NED, the group can be identified by the 'groupid' field in the Proc2KA_Merge table.
- In the v4 NED, a separate field for groups does not exist, but, sites can be identified as having a common server which is identified by the field 'install_no' in the Proc_Main table.
- Trend analysis requires the use of data from "stable" sites. The number of sites which can be considered stable will depend on the interval being studied.

2. Endoscopists

Endoscopists generally perform procedures within only one group although often at more than one site within that group. However, it is possible for providers to perform procedures with more than one group.

For example, a rural endoscopy suite might have two providers each of whom travel from their respective primary practice sites to perform procedures one day each week at the rural site. One of those endoscopists might use the CORI software at their primary site, while the other does not.

Endoscopists are identified by a unique identifier and cannot be reidentified from data in the NED. Additional information about endoscopists include their degree, year of graduation from medical school, when known, and presence and/or year of certification in pertinent medical specialties. These values have been verified when possible. A flag is present in the tables containing provider data which indicate whether or not this independent verification was possible.

Caveats on the use of endoscopist data.

- The endoscopists identified with each procedure are those identified at the time of the procedure as the "responsible endoscopist."
- In addition, at academic medical centers, if a fellow is present for the procedure, the fellow is identified.
- There is some variability in the use of these two fields, however, and who actually performed the procedure when a fellow is present is not certain.
- Although most group practices either use or do not use CORI, if an endoscopist performs procedures at more than one site, the NED may not include all of that endoscopist's procedures.
- Although not all procedures performed by an endoscopist at a site have to be entered into the CORI software, prior quality assurance evaluations have indicated that this is almost always the case.
- Providers who used both CORI v3 and CORI v4 can be identified across the two versions. In the v3 NED, the Physician table contains a column 'physician' which is a unique ID for the responsible physician. In the v4 NED, if there are records from that physician, the same unique ID will be found in the Staff table, 'v3physician' column.
- In spite of this, in the v3 NED, some providers may have more than one identifier. No similar duplication should be present in the v4 NED due to tighter control on participation.
- The primary unit of analysis for the NED is the procedure. However, some studies have been performed with the endoscopist as the unit of analysis, primarily related to quality measures (see reference below).

Reference:

Williams JE, Holub JL, Faigel DO. Polypectomy rate is a valid quality measure for colonoscopy: results from a national endoscopy database. *Gastrointest Endosc.* 2012 Mar;75(3): 576-82. (PMC3839247; PMID: 22341104)

3. Patients

Patients are identified in the NED by a unique identifier and cannot be reidentified from data in the NED. Patients will have different identifiers in v3 and in v4 NED but if they had procedures performed in both versions at the same site, then the v4 identifier is also found in the v3 main table (proc2kaMerge). If patients have procedures at more than one site, they cannot be reliably linked in the v3 NED and cannot be linked at all in the v4 NED. Other information about patients includes their race, ethnicity, gender, and age in years at the time of the procedure.

Caveats on the use of patient data.

- The NED may not have data on all procedures performed on a patient, since a patient may have had procedures performed at endoscopy sites which do not use the CORI software.
- In addition, patients cannot be reliably linked across sites that are not in the same group.

- The primary unit of analysis for the NED is the procedure. However, longitudinal studies have been performed with the patient as the unit of analysis using patients who have had multiple procedures at a CORI site.

Reference:

Lieberman D, Holub JL, Morris CD, Logan J, Williams JL, Carney P. Low rate of Large Polyps (>9 mm) Within 10 Years After an Adequate Baseline Colonoscopy With no Polyps. *Gastroenterology*. 2014 Aug;147(2):343-50. doi: 10.1053/j.gastro.2014.04.020. Epub 2014 Apr 22. (PMC4281886; PMID: 24768680)

4. Procedures

The primary unit of analysis in the NED is the procedure. Figure 3 shows the count of procedures of all types by year of the procedure. The change from v3 to v4 software occurred over several years between 2009 and 2012. Each site/group used only a single version at any one time although a site may have used both versions during their transition year.

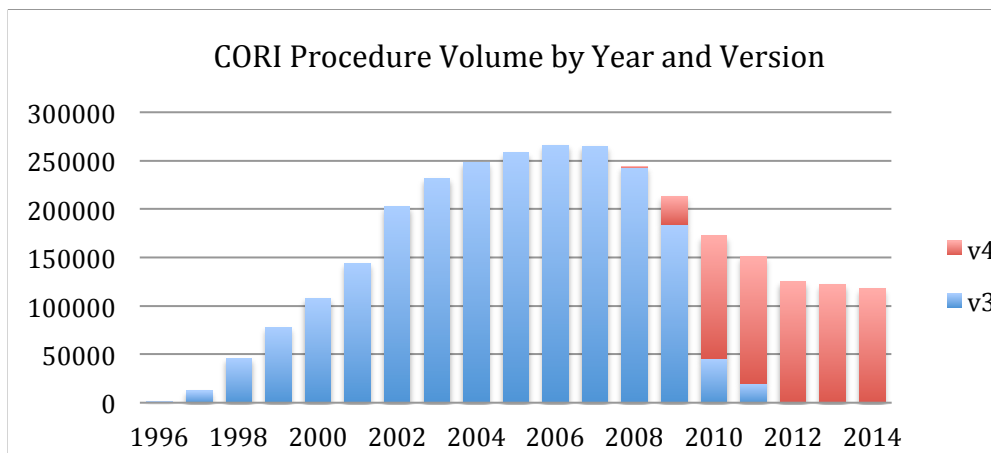


Figure 3. CORI Procedure Volume by Year and Version. V3 was in use from 2000 - 2012. V4 was in use 2008 – 2014. Data prior to 2000 is not considered reliable and is not included in the NED.

Procedure types. Seven types of GI endoscopic procedures are documented in the NED. In addition, bronchoscopies can also be documented as well as non-endoscopic procedures such as liver biopsies. In the v3 NED, both adult and pediatric procedures were documented using the same software modules, so the data from them will appear in the same tables. With the v4 NED, however, pediatric procedure modules became available, although their use for pediatric patients was not required. The data from the pediatric modules can, therefore, be found in the pediatric tables as well as the adult tables

Procedure data. Procedure data was captured using the CORI software in a mix of structured and unstructured fields. Every effort was made to assure data quality through the careful

construction of the CORI software. Whenever possible a structured field was provided for documentation but providers could document in any way that they preferred in v3. For example, in a field requesting polyp size (mm) free text could be entered. Some fields were required but most were not. As such, data in the NED is both "dirtier" and less complete than clinical study data.

General categories of data include:

- History and physical examination documentation
- Liver disease description
- Procedure indications, including the primary indication for a procedure
- Procedure information such as the specific procedure or procedures performed along with information about conduct of the procedure, patient risk, and sedation administered.
- Either a normal examination is documented or a minimum of one finding is documented for each procedure.
- Procedure and immediate post-procedure adverse events
- Assessment and plan formulated at the end of the procedure
- Postprocedure documentation including postprocedure adverse events
- Pathology reports

Caveats on the use of procedure data.

- Data that is captured reliably in the CORI software include the procedure indications, findings, immediate complications, and data on conduct of the procedure.
- History and physical examination fields were not used consistently and have rarely been used for research. In addition, this data is missing for some v3 procedure types.
- A primary indication for a procedure was always required and one or more other indications may be given.
- For colonoscopies, important indications are low and high risk screening and surveillance examinations. These have been defined as needed by specific studies. (see Lieberman reference below)
- While most finding data is felt to be highly reliable, some common findings may not be consistently documented, such as diverticulitis or hemorrhoids found on colonoscopy or flexible sigmoidoscopy procedures.
- There is a general lack of logic checks in CORI, particularly in v3. For example, users could select both "normal exam" and a finding on a colonoscopy procedure.
- Diagnoses associated with each finding are endoscopic diagnoses and not pathologic diagnoses, i.e. the diagnoses given by the endoscopist at the time of the procedure. As noted below, some pathology data is available, but not for all procedures.
- If an anesthesiologist or nurse anesthetist assisted the procedure, the sedation medications may not have been captured in the NED. However, studies have been successfully performed using sedation data available (see Wang reference below).
- ICD codes were provided to sites to aid in billing. However, these were customizable by site. Text fields are available which describe the specific procedure or procedures performed.

- Postprocedure documentation was not consistently performed using the CORI software, so that long term adverse events cannot be determined using the NED (see Enestvedt reference below).

References:

Lieberman DA, Williams JL, Holub JL, Morris CD, Logan JR, Eisen GM, Carney P. Race, Ethnicity, and Sex Affect Risk for Polyps Greater than 9 mm in Average-risk Individuals. *Gastroenterology*. 2014 Aug;147(2):351-8. doi: 10.1053/j.gastro.2014.04.037. Epub 2014 Apr 29. (PMC4121117; PMID: 24786894)

Wang A, Hoda KM, Holub JL, Eisen GM. Does level of sedation impact detection of advanced neoplasia? *Dig Dis Sci*. 2010 Aug;55(8):2337-43. Epub 2010 Apr 22. (PMC3846178; PMID 20411420)

Enestvedt BK, Eisen GM, Holub J, Lieberman DA. Is the American Society of Anesthesiologists classification useful in risk stratification for endoscopic procedures? *Gastrointest Endosc*. 2013 Mar;77(3):464-71. doi: 10.1016/j.gie.2012.11.039 (PMC3816502; PMID: 23410699)

Pathology

Pathology reports or extracts may have been added to the NED by several mechanisms and are only available for a minority of sites. In the v3 NED, pathology had to be added by opening the procedure and then either by pasting or entering text or by entering pathologic diagnoses in structured fields. In the v4 NED, pathology reports could be automatically associated with the procedure using an HL7 interface in addition to one of the above methods. In research done using pathology by CORI staff, the pathology fields were manually examined for categorization of results. While labor intensive, we have found this step necessary in order to properly categorize and analyze pathology data in CORI.

References:

Lieberman D, Moravec M, Holub J, Michaels L, Eisen G. Polyps size and advanced histology in patients undergoing colonoscopy screening: Implications for CT colonography. *Gastroenterology*. 2008 Oct;135(4):1100-5. Epub 2008 Jul 3. (PMCID: PMC2581902)

Diamond SJ, Enestvedt BK, Jiang Z, Holub JL, Gupta M, Lieberman DA, Eisen GM. Adenoma detection rate increases with each decade of life after 50 years of age. *Gastrointest Endosc*. 2011 Jul;74(1):135-40. Epub 2011 May 25. (PMC3878149; PMID: 21612774)

Deidentification of data

The National Endoscopic Databases, as used for research in association with CORI, contain identifiable data (dates of birth and service, addresses to zip code level, and extensive free text fields). The NED data, as found in the NIDDK Repository, has been deidentified. Steps taken to deidentify this data include:

- Truncation of all dates to year of service / event or age in years at the time of the procedure

- Restriction of addresses to the first three digits of the zip code but only in zip codes having a population > 25,000 based on 2010 US Census Population Data.
- Redaction of whole fields known to be at high risk and low utility for research, such as comments fields
- Redaction of patient, provider and facility names and dates from fields which are low risk but high utility for research.