

Balancing Privacy and Utility with Smart Anonymization



Medical History in Patient Narratives



Speaker Introduction



Zach Weingarden, M.S.

Director of AI Technology & Applications
TrialAssure

Master of Science, Biomedical Engineering
University of Michigan

Advantages of Retaining Medical History

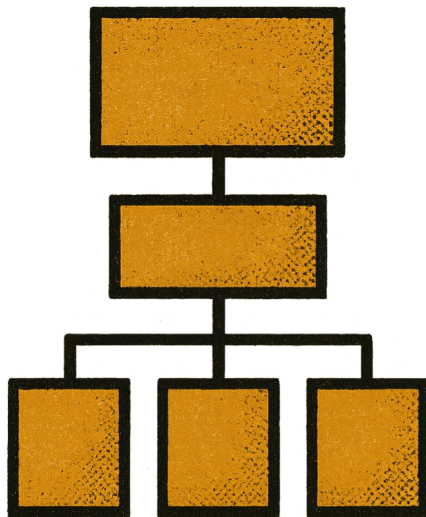
- Improved research validity
- Pattern discovery and comorbidities
- Utility for AI and predictive care
- Over-redaction erodes context and undermines risk assessment

Challenge #1: Patient Privacy

- Medical terms are often very specific and identifying
- Specific years are present
- Patients often have unique combinations of medical background
- Medical background can be well-known to acquaintances



Medical Term Generalization



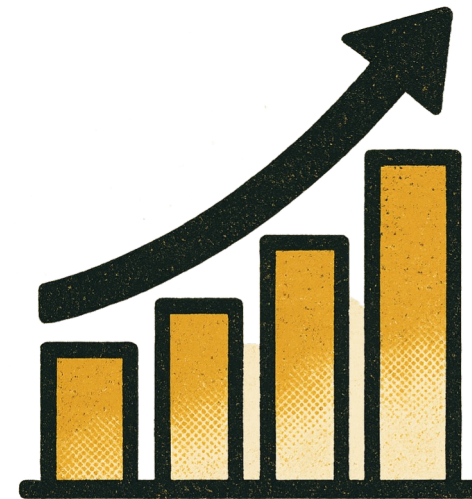
Specificity

MedDRA's hierarchical structure enables selection of broader terms, reducing granularity while preserving clinical meaning.



Privacy

Applying less specific MedDRA terms enhances anonymization, lowering identifiability risk in shared datasets.

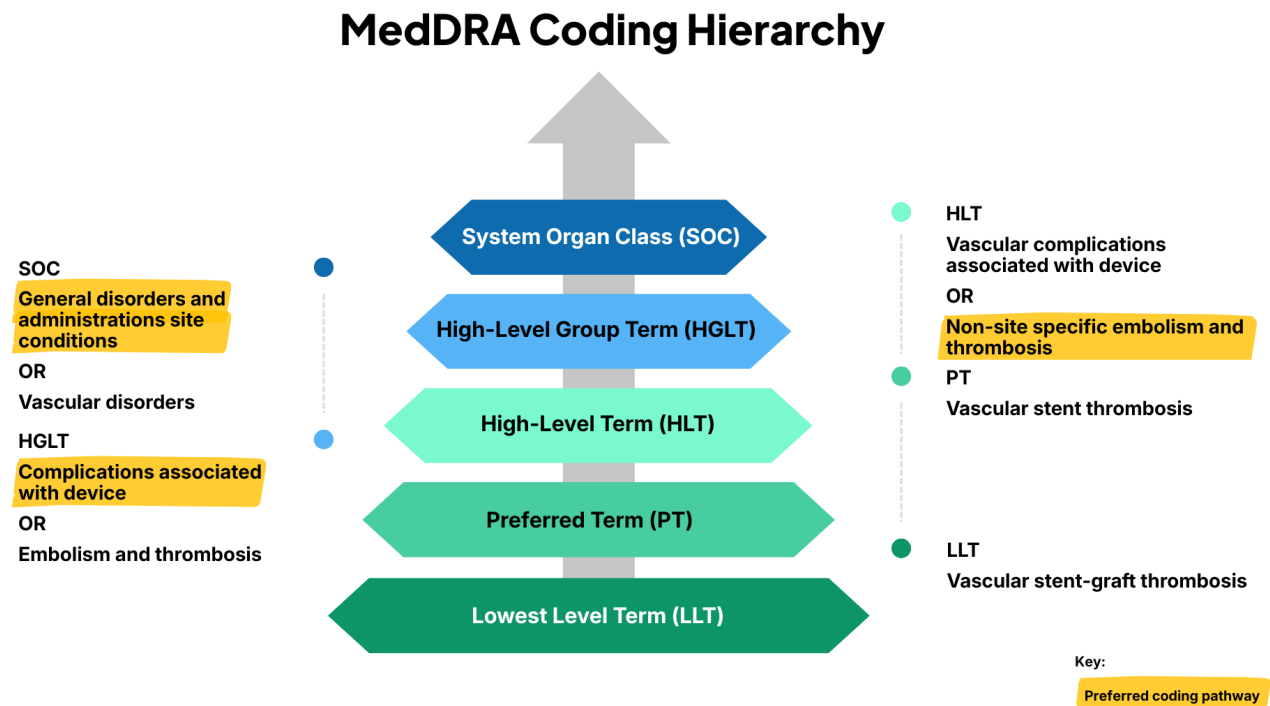


Data Utility

Studies show a 25% increase in data usability metrics when MedDRA-guided generalization is implemented, maintaining valuable insights.

Medical Term Generalization

- Replace specific *lower-level terms* with more general *higher-level terms*
- Preserves meaning
- Problem: higher-level terms are often lengthier than lower-level terms
 - Creates special issues and retaining pagination



Narrative Example

Patient ID: PRM001-01-111

Treatment Group: Placebo

Age at Baseline / Race / Sex: 57 year-old / white / Male

Preferred term(s): Allergic reaction, bronchitis

Patient No. PRM001-01-111, a 57 year-old white male from the United States, treated with the placebo medication, experienced the adverse events of worsening/aggravated allergic reaction to contrast dye, bronchitis. Medical history included hypertension (1977), myocardial infarction (1998), asthma (2002), coronary artery disease (2004), ventricular septal defect repair (2004), coronary artery stent placement (2004), constipation (2008), pulmonary arterial hypertension (2008), gout (2008), hyperlipidemia (2008), allergic rhinitis (2009), tingling in fingers (neuropathy) (2010), nephrotic syndrome (2010), gallbladder stone (2011), heparin allergy (2011), metabolic acidosis (06 Apr 2007), hyperphosphatemia (30 Jul 2008), hypercalcemia (11 Aug 2008), hyperkalemia (17 Oct 2008), ESRD (17 Oct 2008), secondary hyperparathyroidism (24 Oct 2008), anemia (13 Nov 2008), iron deficiency (02 Dec 2008), intra-dialytic hypotension (11 May 2009), depression (04 Jun 2009), pain from hemodialysis needle insertion (02 Jul 2009), intra-dialytic itching (21 Mar 2011 to 17 Aug 2011), renal cell carcinoma (17 Nov 2011 to unknown Dec 2011), pneumonia (30 Dec 2011 to unknown Jan 2012), renal stent insertion (unknown Aug 2011 to 24 Aug 2011), hematuria (Aug 2011), edema (unknown date), Ile (Sept 2011), and hypertension (unknown date). Concomitant medications included Lisinopril, Crestor, Terazosin, Sensipar, Clonidine, Paricalcitol, Gabapentin, Acetaminophen, Renvela, Fosrenol, Metoprolol, Diazepam, influenza virus vaccine, PPD read TST and Paricalcitol.

On 04 Apr 2012, the subject enrolled in the study. On 19 Jun 2012, the patient had the adverse event of "Tingling" in legs, which resolved the same day. No medical action was taking. On 25 Oct 2012 the patient had high central venous pressures and 70% cephalic vein stenosis. The patient underwent a fistulagram, angioplasty, and EKG. He received single dose of 0.5 mL Lidocaine 2%, and intravenous 2 mg versed, and 100 ug fentanyl. These events resolved on the same day. The patient experienced worsening/aggravated allergic reaction to contrast dye on 26 Oct 2012, the dose of medication was not changed and 12.5 mg diphenhydramine was given to alleviate symptoms. This event resolved on the same day. On 02 Nov 2012 the patient experienced bronchitis. Twice daily 500mg Amoxicillin was administered. On 11 Nov 2012, the patient recovered from bronchitis. 13 Nov 2012, the patient stopped taking Amoxicillin. The dosage of placebo was not changed for any of these events. On 28 Dec 2012 the patient ended the study.

The investigator determined that the events (worsening/aggravated allergic reaction to contrast dye, bronchitis, high central venous pressures, "tingling" in legs, 70% cephalic vein stenosis) were not related to the study drug as the patient was in the placebo group.



Medical History

Masked or generalized



Adverse Events

Retained

Challenge #2: Volume

- Hundreds or thousands of narratives
- Requires a repeatable, scalable, and accurate solution

Challenge #3: Separating from AEs

- Keyword-based rules approach can also flag:
 - Adverse events related to the study (which should be retained)
 - Study-level topics, like inclusion criteria
- Requires:
 - Detailed pre-configuration: pages ranges, locations on the page, terms to look for, etc.
 - Meticulous clean-up

Further Refinement

- Apply traditional concepts to improve AI reliability
- Examples:
 - AI agents identify page ranges where individual narratives are present
 - Presence and location of repeated medical history tables
 - Leverage study-specific “clues” – common medical history terms, most frequent adverse events, study date ranges
- Implement smart solutions to space issues

Hybrid Approach

- AI isolates medical history sections from adverse events and study level topics (e.g. inclusion criteria)
- Apply MedDRA-based substitutions based on a traditional algorithmic risk-based de-identification approach

Patient ID: PRM001-11-604

Treatment Group: Placebo

Age at Baseline/Race/Sex: 58-year-old / Black or African American / Female

Preferred term(s): Worsening of fluid overload

Patient No. PRM001-11-604, a 58 year-old Black or African American female from USA, treated with placebo, experienced an SAE of worsening of fluid overload during the study. Medical history included: ex-smoker (1964 to 2003), coronary artery disease (2002), left eye vision loss (2003), left intraocular lens placement (2003), depression (2006 to 2010), spondylosis (2008), arthritis (2008), shortness of breath (2008), hyperlipidemia (2009), hypercholesterolemia (2009), left carotid endarterectomy (2009), hypertension (2009), diabetes type II (2009), diabetic nephropathy (2009), right eye cataract (2009), intermittent neck pain-mild (2009), peripheral neuropathy (2009), peripheral artery disease (2010), hand numbness (2010), hand pain (2010), fluid overload (2010), colon polyp (27 Aug 2008), coronary stent placement (22 Jan 2009), hyperphosphatemia (04 Jan 2010), inter dialytic hypotension (04 Jan 2010), facial edema (04 Jan 2010), lower extremity edema (04 Jan 2010), ESRD (04 Jan 2010), angioplasty-fistula (09 Feb 2010), pericarditis (19 Feb 2010), angioplasty-fistula (23 Feb 2010), bronchitis (21 Apr 2010 to 28 Apr 2010), AV fistula (05 May 2010), tunneled catheter removal (06 May 2010), anemia-ESRD (02 Jul 2010), bronchitis (03 Nov 2010 to 10 Nov 2010), cardiomegaly (19 Apr 2011), angioplasty-fistula (11 Jul 2011), angioplasty-fistula (11 Aug 2011), drug allergy, and hypertension. Concomitant medications at the time of the event were: Heparin, plavix, Simvastatin, renal tab, insulin, vitamin b-12, fish oil, Centrum, Simcor, Lidocaine 1%, Coreg, Lipitor, Lisinopril, Aspirin, Hectoral, Atarax, Zemplar, Pentoxifylline, and oxygen.

On 26 Oct 2011, the patient entered the study and received the first dose of the placebo drug. On 06 Dec 2011, the patient experienced the SAE of worsening of fluid overload and was hospitalized. During this time treatment with placebo was interrupted. The patient experienced non-ST elevation myocardial infarction and underwent cardiac catheterization on 07 Dec 2011. The patient underwent cardiac catheterization on 08 Dec 2011 and subsequently recovered from non-st elevation myocardial infarction. On 09 Dec 2011, the patient started 3.125 mg of twice daily Coreg, 40 mg Lipitor, and 5 mg Lisinopril. The patient recovered from worsening of fluid overload and discharged on 12 Dec 2011. On 16 Dec 2011, the patient was diagnosed with renal osteodystrophy and started 2 ug intravenous Hectoral. The same day, the patient experienced insomnia. 19 Dec 2011, the patient was given 25 mg Atarax for the insomnia. On 02 Jan 2012, 2 ug Hectoral was discontinued. On 03 Jan 2012, Atarax was stopped. The patient had elevated ferritin on 04 Jan 2012, which resolved on 18 Jan 2012. On 26 Jan 2012, the patient had an angiogram of fistula with arteriogram. 27 Feb 2012 the patient experienced an upset stomach. This resolved the same day. 17 Apr 2012, the patient had a chest x-ray. The patient left the study on 29 Jun 2012. The events of renal osteodystrophy and insomnia were ongoing at this time.

The investigator determined that the event (worsening of fluid overload) was not related to the study drug.

Hybrid Approach

Rule-Based Methods

Utilizes predefined patterns and rules to identify sensitive data for anonymization. Ensures compliance with regulatory standards and provides predictable, auditable results across datasets.

AI-Powered Entity Recognition

Leverages machine learning to detect complex entities and relationships, adapting to evolving privacy risks and capturing nuances missed by purely rule-based approaches.

Contextual Analysis & Benefits

Assesses data context to apply balanced granularity, minimizing unnecessary redaction. Hybrid models streamline workflows and reduce information loss.

Thank you!

Any questions?




TrialAssure®