

CRC Model: Feature Pipeline by Notebook

This document traces every feature from raw clinical data through each reduction stage, annotated with FHIR availability.

FHIR Availability Key: - **Yes** – Standard FHIR resource with well-defined codes (LOINC, SNOMED, ICD-10) - **Partial** – FHIR resource exists but implementation and data quality vary significantly across systems - **No** – No standard FHIR mapping; system-specific data

FHIR-Derivable: For engineered features (stages 2-4), a feature is marked “Yes” if it can be computed entirely from FHIR-available raw data.

Book 0: Demographics & Cohort

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource
Gender	Patient table	Yes	Patient.gender
Birth date (for age)	Patient table	Yes	Patient.birthDate
Marital status	Patient table	Yes	Patient.maritalStatus
Race	Patient table	Partial	US Core Patient.race extension
Encounter dates	Encounter tables	Yes	Encounter.period
PCP assignment (with eff/term dates)	pat_pcp table	Partial	CareTeam / PractitionerRole (varies by system)
Months since first encounter	Encounter tables	Yes	Derived from Encounter.period

2. Features Engineered (before reduction)

Feature	Description	FHIR-Derivable
AGE	Age at observation	Yes
IS_FEMALE	Binary gender flag	Yes
IS_MARRIED_PARTNER	Married or partner	Yes
RACE_CAUCASIAN	One-hot race	Partial
RACE_BLACK_OR_AFRICAN_AMERICAN	One-hot race	Partial
RACE_HISPANIC	One-hot race	Partial
RACE_ASIAN	One-hot race	Partial
RACE_OTHER	One-hot race	Partial
OBS_MONTHS_PRIOR	Months in system	Yes
HAS_PCP_AT_END	Active PCP at observation date	Partial
HAS_FULL_24M_HISTORY	24+ months of encounters	Yes
months_since_cohort_entry	Time since first cohort observation	Yes

3. Features Sent to Book 9 (after Book 8 compilation)

Feature	FHIR-Derivable
IS_FEMALE	Yes
IS_MARRIED_PARTNER	Yes
HAS_PCP_AT_END	Partial
months_since_cohort_entry	Yes
RACE_CAUCASIAN	Partial
RACE_BLACK_OR_AFRICAN_AMERICAN	Partial
RACE_HISPANIC	Partial
RACE_ASIAN	Partial
RACE_OTHER	Partial
HAS_FULL_24M_HISTORY	Yes

4. Features in Final Model (40 features, Book 9 output)

Feature	FHIR-Derivable
IS_FEMALE	Yes
IS_MARRIED_PARTNER	Yes
HAS_PCP_AT_END	Partial
months_since_cohort_entry	Yes

Book 1: Vitals

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource / LOINC
Systolic blood pressure	pat_enc_enh	Yes	Observation (85354-9)
Diastolic blood pressure	pat_enc_enh	Yes	Observation (85354-9)
Weight (ounces)	pat_enc_enh	Yes	Observation (29463-7)
Pulse / heart rate	pat_enc_enh	Yes	Observation (8867-4)
BMI	pat_enc_enh	Yes	Observation (39156-5)
Temperature	pat_enc_enh	Yes	Observation (8310-5)
Respiratory rate	pat_enc_enh	Yes	Observation (9279-1)

2. Features Engineered (~50 before reduction)

Feature	Description	FHIR-Derivable
WEIGHT_OZ	Latest weight	Yes
WEIGHT_LB	Latest weight (lbs)	Yes

Feature	Description	FHIR-Derivable
BP_SYSTOLIC	Latest systolic BP	Yes
BP_DIASTOLIC	Latest diastolic BP	Yes
PULSE	Latest heart rate	Yes
BMI	Latest BMI	Yes
TEMPERATURE	Latest body temp	Yes
RESP_RATE	Latest respiratory rate	Yes
DAY_SINCE_WEIGHT	Recency of weight measurement	Yes
DAY_SINCE_SBP	Recency of BP measurement	Yes
DAY_SINCE_PULSE	Recency of pulse measurement	Yes
DAY_SINCE_BMI	Recency of BMI measurement	Yes
DAY_SINCE_TEMPERATURE	Recency of temp measurement	Yes
DAY_SINCE_RESP_RATE	Recency of resp measurement	Yes
WEIGHT_CHANGE_PCT_6M	6-month weight change %	Yes
WEIGHT_CHANGE_PCT_12M	12-month weight change %	Yes
WEIGHT_MEASUREMENT_COUNT_12M	Count of weight measurements	Yes
WEIGHT_VOLATILITY_12M	Weight standard deviation	Yes
WEIGHT_TRAJECTORY_SLOPE	Linear regression slope of weight	Yes
WEIGHT_TRAJECTORY_R2	R-squared of weight trend	Yes
MAX_WEIGHT_LOSS_PCT_60D	Max weight loss in 60 days	Yes
BMI_CHANGE_6M	BMI change over 6 months	Yes
BMI_CHANGE_12M	BMI change over 12 months	Yes
BMI_LOST_OBESE_STATUS	Was obese, now not	Yes
BMI_LOST_OVERWEIGHT_STATUS	Was overweight, now normal	Yes
BP_MEASUREMENT_COUNT_6M	BP measurement count (6mo)	Yes
SBP_VARIABILITY_6M	Systolic BP std dev (6mo)	Yes
DBP_VARIABILITY_6M	Diastolic BP std dev (6mo)	Yes
PULSE_PRESSURE_VARIABILITY_6M	Pulse pressure std dev	Yes
AVG_PULSE_PRESSURE_6M	Average pulse pressure	Yes
PULSE_PRESSURE	SBP minus DBP	Yes
MEAN_ARTERIAL_PRESSURE	$(2*DBP + SBP) / 3$	Yes
WEIGHT_LOSS_5PCT_6M	$\geq 5\%$ weight loss (6mo)	Yes
WEIGHT_LOSS_10PCT_6M	$\geq 10\%$ weight loss (6mo)	Yes
RAPID_WEIGHT_LOSS_FLAG	$>5\%$ loss in 60 days	Yes
HYPERTENSION_FLAG	$SBP \geq 140$ or $DBP \geq 90$	Yes
SEVERE_HYPERTENSION_FLAG	$SBP \geq 160$ or $DBP \geq 100$	Yes
TACHYCARDIA_FLAG	Pulse >100	Yes
UNDERWEIGHT_FLAG	$BMI < 18.5$	Yes
OBESE_FLAG	$BMI \geq 30$	Yes
FEVER_FLAG	Temp >100.4	Yes
TACHYPNEA_FLAG	Resp rate >20	Yes
BRADYPNEA_FLAG	Resp rate <12	Yes
CACHEXIA_RISK_SCORE	BMI + weight loss composite	Yes
weight_loss_severity	Ordinal 0-3	Yes
vital_recency_score	Ordinal 0-3 measurement freshness	Yes

Feature	Description	FHIR-Derivable
cardiovascular_risk	HTN + obesity combined	Yes
abnormal_weight_pattern	Rapid loss or declining trajectory	Yes
bp_instability	SBP variability >15	Yes

3. Features Sent to Book 9 (24 after book-level reduction)

Feature	FHIR-Derivable
WEIGHT_OZ	Yes
BP_SYSTOLIC	Yes
BMI	Yes
PULSE	Yes
PULSE_PRESSURE	Yes
WEIGHT_CHANGE_PCT_6M	Yes
MAX_WEIGHT_LOSS_PCT_60D	Yes
WEIGHT_TRAJECTORY_SLOPE	Yes
WEIGHT_LOSS_10PCT_6M	Yes
RAPID_WEIGHT_LOSS_FLAG	Yes
DAY_SINCE_WEIGHT	Yes
SBP_VARIABILITY_6M	Yes
HYPERTENSION_FLAG	Yes
TACHYCARDIA_FLAG	Yes
FEVER_FLAG	Yes
OBESE_FLAG	Yes
UNDERWEIGHT_FLAG	Yes
CACHEXIA_RISK_SCORE	Yes
weight_loss_severity	Yes
vital_recency_score	Yes
cardiovascular_risk	Yes
abnormal_weight_pattern	Yes
bp_instability	Yes

4. Features in Final Model (8 of the 40)

Feature	FHIR-Derivable
vit_BP_SYSTOLIC	Yes
vit_WEIGHT_OZ	Yes
vit_MAX_WEIGHT_LOSS_PCT_60D	Yes
vit_WEIGHT_CHANGE_PCT_6M	Yes
vit_WEIGHT_TRAJECTORY_SLOPE	Yes
vit_vital_recency_score	Yes

Feature	FHIR-Derivable
vit_RECENCY_WEIGHT	Yes
vit_SBP_VARIABILITY_6M	Yes

Book 2: ICD-10 Diagnoses

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource
Outpatient ICD-10 diagnosis codes	pat_enc_dx_enh	Yes	Condition
Inpatient ICD-10 diagnosis codes	hsp_acct_dx_list_enh	Yes	Condition
Problem list (chronic conditions)	problem_list_hx_enh	Yes	Condition (category=problem-list-item)
Structured family history	family_hx	Partial	FamilyMemberHistory

ICD-10 code families extracted: GI bleeding (K62.5, K92), bowel changes (K59, R19.4), abdominal pain (R10), weight loss (R63.4), fatigue (R53), anemias (D50-D64), polyps (D12, K63.5), IBD (K50-K51), prior malignancy (Z85), diabetes (E10-E11), obesity (E66), family history (Z80, Z83.71), diverticular disease (K57), and 20+ additional categories.

2. Features Engineered (~88-116 before reduction)

Feature Category	Count	Examples	FHIR-Derivable
Symptom flags (12mo)	6	BLEED_FLAG_12MO, PAIN_FLAG_12MO, Yes ANEMIA_FLAG_12MO	Yes
Symptom counts (12mo)	6	BLEED_CNT_12MO, PAIN_CNT_12MO	Yes
Symptom flags (24mo)	6	BLEED_FLAG_24MO, PAIN_FLAG_24MO	Yes
Symptom counts (24mo)	6	BLEED_CNT_24MO, PAIN_CNT_24MO	Yes
Risk factor flags	13	POLYPS_FLAG_EVER, IBD_FLAG_EVER, MALIGNANCY_FLAG_EVER	Yes
Enhanced family history	9	FHX_CRC_COMBINED, FHX_FIRST_DEGREE_CRC, HIGH_RISK_FHX_FLAG	Partial
Comorbidity scores	6	CHARLSON_SCORE_12MO, ELIXHAUSER_SCORE_12MO	Yes
Recency features	3	DAYS_SINCE_LAST_BLEED, DAYS_SINCE_LAST_ANEMIA	Yes
Composite scores	6	CRC_SYMPTOM_TRIAD, SYMPTOM_BURDEN_12MO	Yes
Acceleration features	6	BLEED_ACCELERATION, SYMPTOM_ACCELERATION	Yes
Additional conditions	14	IRON_DEF_ANEMIA_FLAG_12MO, CONSTIPATION_FLAG_12MO	Yes

Feature Category	Count	Examples	FHIR-Derivable
Clinical composites	6	INFLAMMATORY_BURDEN, GI_COMPLEXITY_SCORE	Yes

3. Features Sent to Book 9 (26 after book-level reduction)

Feature	FHIR-Derivable
icd_BLEED_FLAG_12MO	Yes
icd_BLEED_CNT_12MO	Yes
icd_ANEMIA_FLAG_12MO	Yes
icd_IRON_DEF_ANEMIA_FLAG_12MO	Yes
icd_PAIN_FLAG_12MO	Yes
icd_BOWELCHG_FLAG_12MO	Yes
icd_WTLOSS_FLAG_12MO	Yes
icd_MALIGNANCY_FLAG_EVER	Yes
icd_CRC_SYMPTOM_TRIAD	Yes
icd_IDA_WITH_BLEEDING	Yes
icd_SYMPTOM_BURDEN_12MO	Yes
icd_METABOLIC_SYNDROME	Yes
icd_POLYPS_FLAG_EVER	Yes
icd_IBD_FLAG_EVER	Yes
icd_HIGH_RISK_HISTORY	Yes
icd_DIABETES_FLAG_EVER	Yes
icd_OBESITY_FLAG_EVER	Yes
icd_FHX_CRC_COMBINED	Partial
icd_HIGH_RISK_FHX_FLAG	Partial
icd_FHX_FIRST_DEGREE_CRC	Partial
icd_CHARLSON_SCORE_12MO	Yes
icd_ELIXHAUSER_SCORE_12MO	Yes
icd_COMBINED_COMORBIDITY_12MO	Yes
icd_severe_symptom_pattern	Yes
icd_genetic_risk_composite	Partial
icd_chronic_gi_pattern	Yes

4. Features in Final Model (7 of the 40)

Feature	FHIR-Derivable
icd_MALIGNANCY_FLAG_EVER	Yes
icd_CHARLSON_SCORE_12MO	Yes
icd_IRON_DEF_ANEMIA_FLAG_12MO	Yes
icd_ANEMIA_FLAG_12MO	Yes
icd_SYMPTOM_BURDEN_12MO	Yes

Feature	FHIR-Derivable
icd_BLEED_CNT_12MO	Yes
icd_PAIN_FLAG_12MO	Yes

Book 3: Social Factors – SKIPPED

All features excluded during development due to data quality issues. No features from this book enter the pipeline.

Book 4: Labs

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource / LOINC
Hemoglobin	res_components	Yes	Observation (718-7)
Hematocrit	res_components	Yes	Observation (4544-3)
MCV	res_components	Yes	Observation (787-2)
MCH	res_components	Yes	Observation (785-6)
MCHC	res_components	Yes	Observation (786-4)
Platelets	res_components	Yes	Observation (777-3)
Iron	res_components	Yes	Observation (2498-4)
TIBC	res_components	Yes	Observation (2500-7)
Ferritin	res_components	Yes	Observation (2276-4)
Transferrin	res_components	Yes	Observation (3034-6)
CRP	res_components	Yes	Observation (1988-5)
ESR	res_components	Yes	Observation (4537-7)
Albumin	res_components	Yes	Observation (1751-7)
ALT	res_components	Yes	Observation (1742-6)
AST	res_components	Yes	Observation (1920-8)
Alkaline phosphatase	res_components	Yes	Observation (6768-6)
Bilirubin (total)	res_components	Yes	Observation (1975-2)
Bilirubin (direct)	res_components	Yes	Observation (1968-7)
GGT	res_components	Yes	Observation (2324-2)
Total protein	res_components	Yes	Observation (2885-2)
CEA	res_components	Yes	Observation (2039-6)
CA 19-9	res_components	Yes	Observation (24108-3)
CA 125	res_components	Yes	Observation (10334-1)
LDH	res_components	Yes	Observation (2532-0)
HbA1c	res_components	Yes	Observation (4548-4)
LDL	res_components	Yes	Observation (2089-1)
HDL	res_components	Yes	Observation (2085-9)

Raw Data Element	Source	FHIR Available	FHIR Resource / LOINC
Triglycerides	res_components	Yes	Observation (2571-8)
FOBT/FIT (stool)	res_components	Yes	Observation (29771-3)

2. Features Engineered (~93 before reduction)

Feature Category	Count	Examples	FHIR-Derivable
Latest lab values	28	HEMOGLOBIN_VALUE, PLATELETS_VALUE, CEA_VALUE	Yes
Days since lab	4	HEMOGLOBIN_DAYS, CEA_DAYS	Yes
Abnormal flags	6	HEMOGLOBIN_ABNORMAL, CEA_ABNORMAL	Yes
Anemia classification	6	ANEMIA_GRADE, IRON_DEFICIENCY_ANEMIA_FLAG	Yes
CEA trends	11	CEA_VELOCITY_PER_MONTH, CEA_DOUBLED_6MO_FLAG	Yes
FOBT features	6	FOBT_POSITIVE_12MO, FOBT_MULTIPLE_POSITIVE_FLAG	Yes
6-month changes	8	HEMOGLOBIN_6MO_CHANGE, ALBUMIN_6MO_CHANGE	Yes
Velocity features	2	HEMOGLOBIN_VELOCITY_PER_MONTH	Yes
Acceleration features	7	HEMOGLOBIN_ACCELERATING_DECLINE PLATELETS_ACCELERATING_RISE	Yes
Drop/pattern flags	5	HEMOGLOBIN_DROP_10PCT_FLAG, THROMBOCYTOSIS_FLAG	Yes
Calculated ratios	7	ALT_AST_RATIO, DE_RITIS_RATIO, ANEMIA_SEVERITY_SCORE	Yes

3. Features Sent to Book 9 (31 after book-level reduction)

Feature	FHIR-Derivable
lab_HEMOGLOBIN_VALUE	Yes
lab_HEMOGLOBIN_DROP_10PCT_FLAG	Yes
lab_HEMOGLOBIN_ACCELERATING_DECLINE	Yes
lab_PLATELETS_VALUE	Yes
lab_PLATELETS_ACCELERATING_RISE	Yes
lab_THROMBOCYTOSIS_FLAG	Yes
lab_IRON_DEFICIENCY_ANEMIA_FLAG	Yes
lab_ANEMIA_SEVERITY_SCORE	Yes
lab_CEA_ELEVATED_FLAG	Yes
lab_FOBT_POSITIVE_12MO	Yes
lab_ALBUMIN_DROP_15PCT_FLAG	Yes
lab_ALBUMIN_VALUE	Yes

Feature	FHIR-Derivable
lab_AST_VALUE	Yes
lab_ALT_VALUE	Yes
lab_ALK_PHOS_VALUE	Yes
lab_CEA_VALUE	Yes
lab_ALT_AST_RATIO	Yes
lab_FERRITIN_VALUE	Yes
lab_CRP_VALUE	Yes
lab_ESR_VALUE	Yes
lab_IRON_VALUE	Yes
lab_HGB_TRAJECTORY	Yes
lab_HEMOGLOBIN_6MO_CHANGE	Yes
lab_HEMOGLOBIN_VOLATILITY	Yes
lab_PLATELETS_RISING_PATTERN_FLAG	Yes
lab_ANEMIA_GRADE	Yes
lab_comprehensive_iron_deficiency	Yes
lab_metabolic_dysfunction	Yes
lab_inflammatory_burden	Yes
lab_progressive_anemia	Yes
lab_any_tumor_marker	Yes

4. Features in Final Model (11 of the 40)

Feature	FHIR-Derivable
lab_HEMOGLOBIN_ACCELERATING_DECLINE	Yes
lab_PLATELETS_ACCELERATING_RISE	Yes
lab_ALT_AST_RATIO	Yes
lab_comprehensive_iron_deficiency	Yes
lab_PLATELETS_VALUE	Yes
lab_THROMBOCYTOSIS_FLAG	Yes
lab_AST_VALUE	Yes
lab_CEA_VALUE	Yes
lab_ALK_PHOS_VALUE	Yes
lab_HEMOGLOBIN_VALUE	Yes
lab_ALBUMIN_VALUE	Yes

Book 5.1: Outpatient Medications

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource
Outpatient medication orders (16 categories)	ORDER_MED_ENH	Yes	MedicationRequest

Medication categories: Iron supplementation, PPIs, NSAIDs/ASA, statins, metformin, laxatives, antidiarrheals, antispasmodics, B12/folate, IBD medications, hemorrhoid/rectal meds, GI bleeding meds, chronic opioids, broad-spectrum antibiotics, hormone therapy, chemotherapy agents.

2. Features Engineered (48 before reduction)

For each of 16 medication categories, 3 features are created (flag, days_since, count_2yr). All FHIR-derivable since they come from MedicationRequest data.

Feature Pattern	Count	FHIR-Derivable
{category}_use_flag	16	Yes
{category}_use_days_since	16	Yes
{category}_use_count_2yr	16	Yes

3. Features Sent to Book 9 (19 after book-level reduction)

Feature	FHIR-Derivable
hemorrhoid_meds_flag	Yes
hemorrhoid_meds_days_since	Yes
iron_use_flag	Yes
laxative_use_flag	Yes
antidiarrheal_use_flag	Yes
ppi_use_flag	Yes
statin_use_flag	Yes
metformin_use_flag	Yes
ibd_meds_days_since	Yes
gi_bleed_meds_days_since	Yes
opioid_use_days_since	Yes
broad_abx_days_since	Yes
hormone_therapy_days_since	Yes
antispasmodic_use_days_since	Yes
nsaid_asa_use_days_since	Yes
gi_symptom_meds (composite)	Yes
alternating_bowel (composite)	Yes
gi_bleeding_pattern (composite)	Yes
hemorrhoid_risk_score (composite)	Yes

4. Features in Final Model

None. All 19 outpatient medication features were eliminated during Book 9 iterative SHAP winnowing.

Book 5.2: Inpatient Medications

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource
Inpatient medication administration records (16 categories)	MAR (Medication Administration Record)	Yes	MedicationAdministration

Same 16 medication categories as outpatient, but from confirmed inpatient administration rather than prescriptions.

2. Features Engineered (48 before reduction)

Same structure as outpatient (16 categories x 3 feature types), all prefixed with `inp_`. All FHIR-derivable.

3. Features Sent to Book 9 (20 after book-level reduction)

Feature	FHIR-Derivable
inp_gi_bleed_meds_flag	Yes
inp_gi_bleed_meds_days_since	Yes
inp_iron_use_flag	Yes
inp_laxative_use_flag	Yes
inp_opioid_use_flag	Yes
inp_hemorrhoid_meds_flag	Yes
inp_hemorrhoid_meds_days_since	Yes
inp_ppi_use_flag	Yes
inp_broad_abx_flag	Yes
inp_antidiarrheal_use_flag	Yes
inp_statin_use_days_since	Yes
inp_metformin_use_days_since	Yes
inp_nsaid_asa_use_days_since	Yes
inp_ibd_meds_days_since	Yes
inp_hormone_therapy_days_since	Yes
inp_acute_gi_bleeding (composite)	Yes
inp_obstruction_pattern (composite)	Yes
inp_severe_infection (composite)	Yes
inp_any_hospitalization (composite)	Yes
inp_gi_hospitalization (composite)	Yes

4. Features in Final Model

None. All 20 inpatient medication features were eliminated during Book 9 iterative SHAP winnowing.

Book 6: Visit History

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource
Outpatient encounter dates & status	PAT_ENC_ENH	Yes	Encounter
Appointment status (completed/no-show)	PAT_ENC_ENH	Yes	Encounter.status
Provider specialty	clarity_ser_enh	Yes	PractitionerRole.specialty
ED encounters	PAT_ENC_HSP_HAR_ENH	Yes	Encounter (class=emergency)
Inpatient admissions & LOS	PAT_ENC_HSP_HAR_ENH	Yes	Encounter (class=inpatient)
GI symptom diagnoses on encounters	pat_enc_dx_enh / hsp_acct_dx_list_enh	Yes	Condition (linked via Encounter)

2. Features Engineered (~34 before reduction)

Feature Category	Count	Examples	FHIR-Derivable
ED counts	4	ED_LAST_90_DAYS, ED_LAST_12_MONTHS, GI_ED_LAST_12_MONTHS	Yes
Inpatient counts	4	INP_LAST_12_MONTHS, GI_INP_LAST_12_MONTHS, TOTAL_INPATIENT_DAYS_12MO	Yes
Outpatient counts	4	OUTPATIENT_VISITS_12MO, PCP_VISITS_12MO, GI_SYMPTOM_OP_VISITS_12MO	Yes
No-shows	1	NO_SHOWS_12MO	Yes
Recency	3	days_since_last_ed, days_since_last_gi	Yes
Flags	7	frequent_ed_user_flag, gi_specialty_engagement_flag	Yes
Composite scores	3	healthcare_intensity_score, primary_care_continuity_ratio	Yes
Reduction composites	5	gi_symptoms_no_specialist, acute_care_reliance	Yes

3. Features Sent to Book 9 (23 after book-level reduction)

Feature	FHIR-Derivable
visit_total_gi_symptom_visits_12mo	Yes
visit_gi_visits_12mo	Yes
visit_gi_ed_last_12_months	Yes
visit_gi_symptom_op_visits_12mo	Yes
visit_gi_specialty_engagement_flag	Yes
visit_gi_symptoms_no_specialist	Yes
visit_ed_last_12_months	Yes

Feature	FHIR-Derivable
visit_ed_last_90_days	Yes
visit_frequent_ed_user_flag	Yes
visit_recent_ed_use_flag	Yes
visit_inp_last_12_months	Yes
visit_recent_hospitalization_flag	Yes
visit_total_inpatient_days_12mo	Yes
visit_outpatient_visits_12mo	Yes
visit_pcp_visits_12mo	Yes
visit_engaged_primary_care_flag	Yes
visit_no_shows_12mo	Yes
visit_healthcare_intensity_score	Yes
visit_acute_care_reliance	Yes
visit_complexity_category	Yes
visit_recent_acute_care	Yes
visit_frequent_ed_no_pcp	Yes
visit_days_since_last_gi	Yes

4. Features in Final Model (7 of the 40)

Feature	FHIR-Derivable
vis_visit_recency_last_gi	Yes
vis_visit_pcp_visits_12mo	Yes
vis_visit_outpatient_visits_12mo	Yes
vis_visit_no_shows_12mo	Yes
vis_visit_gi_symptom_op_visits_12mo	Yes
vis_visit_total_gi_symptom_visits_12mo	Yes
vis_visit_gi_symptoms_no_specialist	Yes
vis_visit_acute_care_reliance	Yes

Book 7: Procedures

1. Raw Data Gathered

Raw Data Element	Source	FHIR Available	FHIR Resource
CT abdomen/pelvis (18 internal codes)	order_proc_enh	Yes	Procedure / ImagingStudy
MRI abdomen/pelvis (9 internal codes)	order_proc_enh	Yes	Procedure / ImagingStudy
Upper GI endoscopy (EGD) (4 codes)	order_proc_enh	Yes	Procedure
Blood transfusion (4 codes)	order_proc_enh	Yes	Procedure
Anoscopy (1 code)	order_proc_enh	Yes	Procedure

Raw Data Element	Source	FHIR Available	FHIR Resource
Hemorrhoid procedures (2 codes)	order_proc_enh	Yes	Procedure
Iron infusions (IV iron formulations)	mar_admin_info_enh	Yes	MedicationAdministration

Note: Colonoscopy deliberately excluded (screened patients already removed from cohort).

2. Features Engineered (~33 before reduction)

Feature Category	Count	Examples	FHIR-Derivable
Procedure counts (12mo)	7	ct_abd_pelvis_count_12mo, iron_infusions_12mo	Yes
Procedure counts (24mo)	7	ct_abd_pelvis_count_24mo, blood_transfusion_count_24mo	Yes
Recency	6	days_since_last_ct, days_since_last_transfusion	Yes
Binary flags	7	high_imaging_intensity_flag, transfusion_history_flag	Yes
Composite scores	6	procedure_intensity_count, comprehensive_gi_workup_flag	Yes

3. Features Sent to Book 9 (17 after book-level reduction)

Feature	FHIR-Derivable
proc_severe_anemia_treatment_flag	Yes
proc_blood_transfusion_count_12mo	Yes
proc_transfusion_history_flag	Yes
proc_iron_infusions_12mo	Yes
proc_iron_infusion_flag	Yes
proc_anemia_treatment_intensity	Yes
proc_acute_bleeding_pattern	Yes
proc_total_imaging_count_12mo	Yes
proc_ct_abd_pelvis_count_12mo	Yes
proc_mri_abd_pelvis_count_12mo	Yes
proc_high_imaging_intensity_flag	Yes
proc_diagnostic_cascade	Yes
proc_upper_gi_count_12mo	Yes
proc_comprehensive_gi_workup_flag	Yes
proc_procedure_intensity_count	Yes
proc_recent_diagnostic_activity_flag	Yes
proc_anal_pathology_flag	Yes

4. Features in Final Model (2 of the 40)

Feature	FHIR-Derivable
proc_total_imaging_count_12mo	Yes
proc_ct_abd_pelvis_count_12mo	Yes

Pipeline Summary

Feature Counts at Each Stage

Book	Domain	Raw Data	Engineered	After Book Reduction	Sent to Book 9	In Final Model
0	Demographics	7	12	—	~10	4
1	Vitals	7	~50	24	24	8
2	ICD-10 Diagnoses	~50 code families	~88-116	26	26	7
3	Social Factors	—	—	—	—	—
4	Labs	29 lab components	~93	31	31	11
5.1	Outpatient Meds	16 med categories	48	19	19	0
5.2	Inpatient Meds	16 med categories	48	20	20	0
6	Visit History	6 encounter types	~34	23	23	8
7	Procedures	7 procedure types	~33	17	17	2
Total			~400+	~172	~172	40

FHIR Availability Summary

Stage	Total Features	FHIR-Derivable	Partial	Not Available
Raw data gathered	~90 elements	~85	~5 (PCP, family hx, race)	0
Sent to Book 9	~172	~165	~7 (PCP, family hx, race features)	0
Final 40 features	40	39	1 (HAS_PCP_AT_END)	0

Key finding: 39 of the 40 final model features can be fully derived from standard FHIR resources. The sole exception is HAS_PCP_AT_END (active PCP at observation date), which requires system-specific PCP assignment data with effective/termination dates – available through the FHIR CareTeam resource in some implementations but not standardized across all systems.

All medication features (39 total from Books 5.1 and 5.2) were eliminated during Book 9 feature selection despite being FHIR-available, meaning the model relies entirely on vitals, labs, diagnoses, visit patterns, and imaging procedures – all of which have strong FHIR coverage.