

Clarity SQL Safety & Data Validation Checklist

Pre-Query Safety | ETL Validation | Clinical Accuracy | TechFitFlow.com

Complete Section 1 before running any new Clarity query. Complete Sections 2-4 before delivering results to any stakeholder. Never skip either section because the query looks simple.

SECTION 1: PRE-QUERY SAFETY CHECKS (COMPLETE BEFORE RUNNING)

	Check Item	Analyst	Done
[]	Read-only access confirmed - Clarity connection is SELECT-only with no DML permissions	_____	___
[]	Query purpose documented - business justification for accessing this patient data defined	_____	___
[]	HIPAA minimum necessary review: only columns needed for the analysis are in the SELECT list	_____	___
[]	No SELECT * used - specific column names listed in SELECT clause	_____	___
[]	Date filter added - WHERE clause includes an indexed date column with a specific range	_____	___
[]	Date range is not wider than the business question requires	_____	___
[]	Half-open interval used for datetime filter (>= start AND < end), not BETWEEN	_____	___
[]	TOP clause added (TOP 100 or TOP 1000) - will remove only after logic is validated	_____	___
[]	All JOIN types are explicit (INNER JOIN or LEFT JOIN) - no implicit comma joins in FROM	_____	___
[]	ZC_ lookup tables joined for all categorical columns in SELECT list	_____	___
[]	No DML statements (INSERT, UPDATE, DELETE, TRUNCATE, DROP) in query	_____	___
[]	If query covers > 90 days of data: will run during off-peak hours (avoid 7-10 AM)	_____	___
[]	Output location confirmed as HIPAA-covered storage (not personal desktop or email)	_____	___

SECTION 2: QUERY LOGIC VALIDATION (REVIEW TOP 100 SAMPLE RESULTS)

	Check Item	Analyst	Done
[]	PAT_ENC_CSN_ID values are distinct - no fan-out from multi-row join without aggregation	_____	___
[]	Patient count uses COUNT(DISTINCT PAT_ID), not COUNT(CSN)	_____	___

	Check Item	Analyst	Done
[]	LOS calculated with DATEDIFF(MINUTE,...)/1440.0, not DATEDIFF(DAY,...)	_____	___
[]	If PAT_ENC_DX joined: PRIMARY_DX_YN = Y filter applied - encounters not duplicated per diagnosis	_____	___
[]	If ORDER_RESULTS joined: COMPONENT_ID filter applied for specific test values	_____	___
[]	If MAR_ADMIN_INFO joined: MAR_ACTION_C filter applied (1=Given where appropriate)	_____	___
[]	Date range returns expected number of records (not zero, not millions)	_____	___
[]	Department names are readable - ZC_ join worked correctly	_____	___
[]	Categorical codes are resolved to names - no raw numeric codes in output	_____	___
[]	NULL values handled explicitly: COALESCE, ISNULL, or documented in query notes	_____	___
[]	Open encounters handled: HOSP_DISCHRG_TIME IS NOT NULL applied if excluding current admits	_____	___

SECTION 3: DATA VALIDATION - ETL RECONCILIATION

	Check Item	Analyst	Done
[]	Record count compared against Epic operational report for same date range and encounter type	_____	___
[]	Count difference within acceptable tolerance (less than 0.1% for most reports)	_____	___
[]	Any difference > 0.1% investigated and root cause documented before delivery	_____	___
[]	Field-level spot check: 10 random records compared between Clarity and Epic application	_____	___
[]	Spot check confirmed: MRN, admission date, discharge date, attending provider, primary diagnosis	_____	___
[]	ETL timestamp checked: Clarity extract ran since the most recent relevant clinical events	_____	___
[]	Near-real-time extract confirmed if report requires same-day data	_____	___

SECTION 4: CLINICAL ACCURACY AND RESULT DELIVERY

	Check Item	Analyst	Done
[]	Clinical SME has reviewed a sample of results for clinical plausibility	_____	___

