

# Sanofi Narrative QC tool

Most of the scenarios are inherited from the Narrative QC tool which was built for Novartis by Analytics team. They were further modified as per Sanofi's requirements specification.

## Scenario 1: Gender Mismatch (known gender)

**If known gender (male/female) is detected in narrative, any usage of opposite gender based pronouns should be captured by tool.**

**E.g.** If the narrative is about a "Male" patient usage of gender-based pronouns like "she, her, herself ..." is an error and should be captured by tool.

**Note:** No exception scenarios provided. If patient's gender info is not mentioned within the narrative, tool will assume as unknown gender.

## Scenario 2: Wrong Pronoun

**Narration is expected to be in indirect speech. Hence usage of 1<sup>st</sup>/ 2<sup>nd</sup> person pronouns is an error.**

**E.g.** 1<sup>st</sup> person pronouns (I, we, us, me ...), 2<sup>nd</sup> person pronouns (you, your, yours, yourself ...) should be captured by tool.

**Note:** No exception scenarios provided. List of pronouns were taken as per general English conventions. Tool doesn't check/ change the narration from direct to indirect speech.

## Scenario 3: Tense error

**The whole narrative should be written in Past tense.**

**E.g.** This case **involves** a 61-year-old male patient. Here "involves" should be captured as error and replaced by its past verb-form "involved".

**Note:** No exception rules provided. Cases where tense is captured as-is from source document (CT cases), are out of scope.

## Scenario 4: Spelling/ No-Singe-Spacing error

Report all spell/ no-spacing errors and provide multiple suggestions, if applicable. Tool should also provide “Update Dictionary” option to QC tool users. In case of multiple suggestions, user should have the flexibility to edit those suggestions.

E.g. **Shewas** admitted to hospital. Highlight and replace “Shewas” by “She was”.  
The **patint** had history of chronic kidney disease. Highlight and replace “patint” by “patient”.

**Note:** Sometimes author names get captured as wrong spelling and their suggestion is suppressed as well as “(Proper Noun)” is mentioned in the error type.

**Case-sensitive spell check** is applied for **Brand & generic** names.

**Considered dictionaries/ libraries:** Internal dictionary (hunspell-R), medDRA 9PT & LLT), FDA (Brand Manufacturer & generic names), Sanofi\_Product\_List\_May20 (as generic names, i.e. in lower case).

## Scenario 5: Special Character

a) Highlight all disallowed special characters.

List of **allowed** special characters in Sanofi is as follows: **!\#\$%&()\*+,-./:;<=>?@[ ]^\_`{|}~**

b) In addition, sentence construction errors for allowed special characters (no/ extra spacing associated with it).

E.g. i) Patient program name: — **“ABC\_0123\_XYZ”**. Disallowed chars highlighted.  
E.g. ii) sentence ends here.**new one start here.** Tool replaces “.n” by “. N”.

**Note:** Single quote is disallowed but it is allowed as apostrophe.  
Default replacement suggestion of disallowed special characters is NULL.

## Scenario 6: Repetition

Highlight repetitive information.

E.g. No relevant history **was received was received.**

**Note:** Currently **consecutive repetitions (of words/ phrases)** are only captured by tool.  
Contextual repetition are out of scope.

### Scenario 7: Extra Spacing

Highlight extra spacing between words.

**Note:** No exception rules.

### Scenario 8: Unknown Gender

If no gender information is reported (or unknown gender is reported), then any male/ female gender based pronouns should be highlighted.

**E.g.** This case refers to a 61 years old patient. He/she had a history of chronic kidney disease. No concurrent conditions were reported for him.

**Note:** No exception rules. Usage of gender-based pronouns as a pair (He/She, him/her, himself/herself ...) will also be detected as error. “The patient” is given as suggestion to these pronouns.

### Scenario 9: Generic Drug (generic-Brand scenario): Out of Scope

### Scenario 10: Date Chronology

The events should be in chronological order, i.e. event dates should appear as from older to recent.

For literature cases, first paragraph will be excluded from chronology check. In literature case, the word “literature” is present within narrative.

**E.g.**

**Note:** No exceptions rule provided for any date exclusion apart from the literature case narrative. All dates including IRD, CSD, expiry dates ... are into consideration for Chronology check.

To compare incomplete dates, ‘01’ and ‘Jan’ are added respectively, where day or month is missing. For e.g. ‘2019’ is treated as “1-Jan-2019”, ‘Feb-2019’ is treated as “1-Feb-2019”

## Scenario 11: Wrong Date Patterns

**Check different date patterns. Find incorrect/ incomplete dates where day/ month is wrong or there exists incomplete year.**

**E.g.** 1-Jna-2020 (wrong month, suggestion: 1-Jan-2020), Jan-20 (incomplete year), 14-13-20 (incomplete year, invalid month) ....

**Note:** This is an additional scenario. Almost all the date patterns are considered.

For MMDDYYYY/ DDMMYYYY pattern dates, first mm-dd-yyyy is checked, then dd-mm-yyyy pattern is checked, if both are invalid, then it is mentioned is error.

## Scenario 12: Abbreviation

**Expansion of abbreviations should be followed by the abbreviation within parenthesis.**

**Proper convention:** *Chronic Kidney Disease (CKD)*

**Any exception of this convention should be reported as an error.**

**Highlight only for first instance.**

**E.g.** Chronic Kidney Disease (CKD something else here) or CKD (Chronic Kidney Disease) or (CKD) Chronic Kidney Disease or COPD – all these cases will be highlighted as error.

**Note:** No expansion-abbreviation cross check is done, as tool doesn't have list of abbreviation & expansions. It is checked based on whether each abbreviation characters are expanded or not.

For **abbreviations written in lower case**, tool **doesn't count it as abbreviation**. For e.g.

Chronic Kidney Disease (ckd)/ ckd/ ckd (Chronic Kidney Disease) are not under this scenario.

**Exceptions:** For these cases expansion is not required. 1. VACCINE NAMES (list shared), 2. some most commonly used allowed abbreviations (e.g. HCP, DOI, IME ...)

## Scenario 13: Dosage Frequency Abbreviation

**Any usage of Dosage Frequency Abbreviations is error (irrespective of whether expansion given or not)**

**E.g.** BID or twice daily (BID) or BID (twice daily) – all these cases will be highlighted as error.

**Note:** No exception cases.

## **Scenario 14: Additional Grammar Scenarios:   Partially Implemented**

- a) Subject vs Auxiliary Verb agreement – singularity/ plurality check. (WIP)**  
e.g. Medical histories **was** not reported. (plural: were)
  - b) Auxiliary Verb vs Main Verb compatibility check. (WIP)**  
e.g. The dose was not **change**. (v3: changed)
  - c) Article related errors (a, an, the), including article for numbers. (Done)**  
e.g. **an** serious (a), **a** 81-years-old patient (an), **a** patient (the)...
-