**Pfizer Generative AI Proof of Technology:**

**International Drug Product Labeling**

Table of Contents

[1 Introduction 2](#_Toc161930730)

[2 Objectives 2](#_Toc161930731)

[3 Scope 2](#_Toc161930732)

[3.1 In-scope 2](#_Toc161930733)

[3.2 Out of scope 2](#_Toc161930734)

[4 Background 2](#_Toc161930735)

[4.1 Pfizer International Labeling Group (ILG) 2](#_Toc161930736)

[4.2 Drug Product Label 3](#_Toc161930737)

[4.3 Current and Future Labeling Workflow 3](#_Toc161930738)

[4.4 Draft Roadmap 4](#_Toc161930739)

[5 Test Scenarios 5](#_Toc161930740)

[5.1 Test Scenario #1 – Generate a new PIL 5](#_Toc161930741)

[5.2 Test Scenario #2 - Update existing PIL (modify, add or subtract content) 6](#_Toc161930742)

[6 Digital Section 7](#_Toc161930743)

[7 Success Criteria 7](#_Toc161930744)

[8 Timelines & Phase gate plan 7](#_Toc161930745)

[Appendices 9](#_Toc161930746)

[Appendix 1 – Sample SmPC to PIL content map 9](#_Toc161930747)

[Appendix 2 – SmPC to PIL conversion examples 11](#_Toc161930748)

# Introduction

Pfizer’s International Labeling Group invites you to participate in a **3-week** Proof of Technology (PoT) hackathon. The objective of this exercise is to help Pfizer identify a strategic partner to build a Generative AI solution dedicated to solving business challenges related to international drug product labeling.

Participants have **3 weeks**, at the end of which, they are expected to:

1. present a live demo to the Pfizer team that demonstrates the ability to achieve the objectives described below. (
2. demonstrate that the developed assets in the Proof of Technology can be easily transferred to the Beta MVP production and deployed upon successful selection;
3. submit a documented production proposal with proposed solution architecture, timelines and total cost of ownership.

This PoT is non-binding and solely for the purpose of identifying the best proposal, at no cost to Pfizer.

# Objectives

The objective of this Proof of Technology Hackathon is to deliver a Gen AI-based solution which meets the quality and accuracy parameters for patient labeling laid out in Test Scenarios 1 and 2 (Refer to section 5). The solution needs to be built on Pfizer's pre-qualified technology (Refer to section 6).

The overall approach also needs to be built with the future in mind, scalable to other labeling use cases, cost effective on an ongoing basis, and incorporate machine learning characteristics to leverage Pfizer's large global volume throughput. The solution also needs to be adaptable to changes in regulation (e.g., new or updated drug product label templates) or technology (e.g., advancements in Gen AI).

# Scope

## In-scope

The following are in-scope for this exercise. The in-scope list of documents, activities and requirements will be expanded in subsequent phases of the project.

* European Quality Review Document (QRD) template
* Summary of Product Characteristics (SmPC) in English
* European Patient Information Leaflets (PIL) in English
* Pfizer qualified LLMs (i.e., ChatGPT v4 and Claude v2)

## Out of scope

The following are only out of scope for this exercise. They will be brought in scope for subsequent phases of the project.

* Patient Information Leaflets in languages other than English
* Non-European drug product label templates

# Background

## Pfizer International Labeling Group (ILG)

Pfizer’s International Labeling Group (ILG) has oversight responsibility for creating and maintaining drug product label documents for Asia Pacific, Europe, Middle East, Africa, and the Americas (excluding the USA).

These responsibilities include the management of ~11,000 label documents and 10,000 to 20,000 label updates per year internationally across Pfizer’s drug portfolio.

Through these labeling documents, we ensure that the information relating to the safety and efficacy of our products reaches patients and healthcare professionals.

## Drug Product Label

The drug product label is a pivotal source of regulated and scientifically validated information that assists healthcare professionals in prescribing and dispensing the medicine and informs consumers about its safe and effective use.

Each national health authority requires drug label information be organized and presented in standardized templates. The templates and the contents of the templates are regulated by the national health authorities and require authorization before being made public.

For example, the European Medicines Agency’s (EMA) template is called the Quality Review Document (QRD)[[1]](#footnote-2) and it consists of the following three sub-templates:

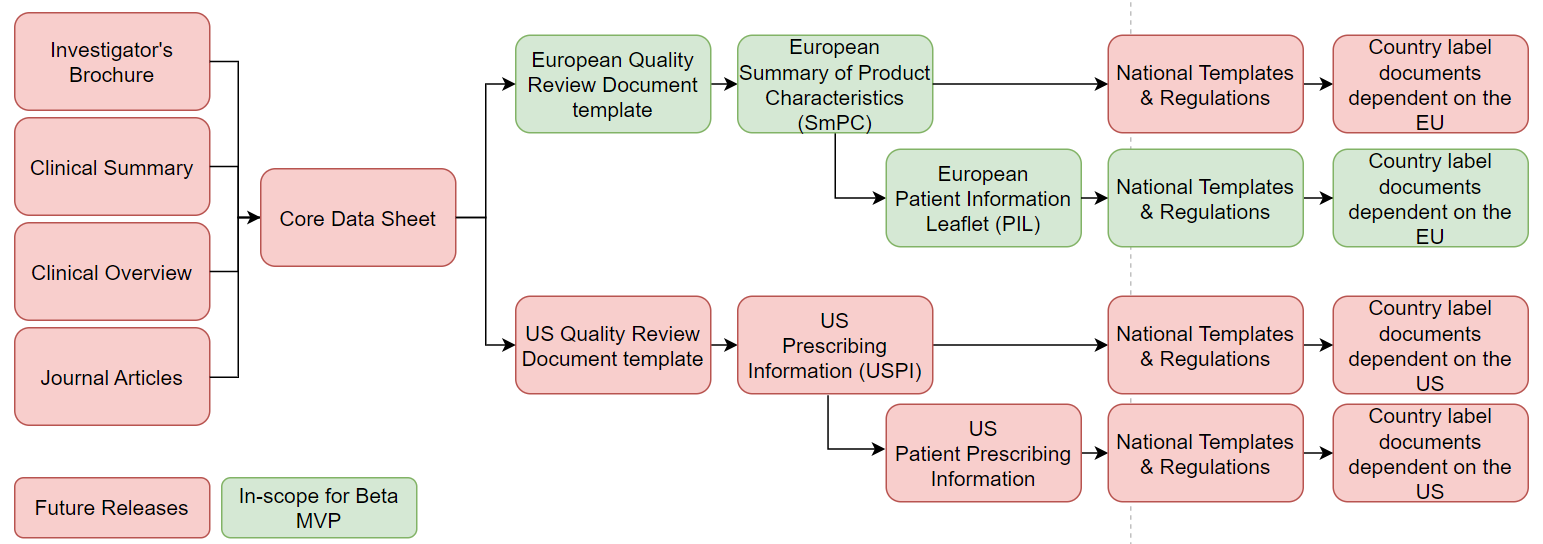
1. Summary of Product Characteristics (SmPC) – captures drug product information written for healthcare professionals
2. Package Labeling – captures drug product information written for patients at a grade 5-8 reading level
3. Package Leaflet – captures drug product information authorized to be present on the drug product’s inner and outer packaging

Once authorized, the label documents are considered public and are posted online on the company website (e.g., [Product List | Pfizer](https://www.pfizer.com/products/product-list)); on national health authority websites (e.g., [Xeljanz | European Medicines Agency](https://www.ema.europa.eu/en/medicines/human/EPAR/xeljanz)); or the content is reused in part or in whole by third party aggregator sites (e.g., [DailyMed](https://dailymed.nlm.nih.gov/dailymed/) or [DrugBank](https://go.drugbank.com/)).

## Current and Future Labeling Workflow

The current paradigm is built around a series of sequential stage gates and dependencies (See Figure 1). The labeling process tracks status and outcomes, where each step relies on the last. It is not predictive and does not leverage precedent. It is manual and resource intensive to build and maintain labels.

**Figure 1 Overview of the end to end workflow used to author drug product label documents**



The portfolio is growing and the demand for digital health and personalisation is increasing. Current processes are not scalable. By applying Gen AI to the labeling process, our ultimate goal is to generate every label of every type everywhere in the world in real time as clinical development proceeds.

Through a series of staged releases, the Beta MVP is the first step on a pathway to full operating capability designed to prove critical concepts, develop reusable foundational components, and deliver incremental value, transforming productivity, quality and timeliness across the labeling lifecycle.

Integration with upstream source document production will enable an “everything everywhere all at once” paradigm, giving unprecedented visibility to commercial brand teams, medical, regulatory, safety and manufacturing colleagues, enabling market preparedness.

The technology foundations will set the agenda for responsible cutting-edge Generative AI in Labeling while delivering modern, digital content to healthcare systems.

## Draft Roadmap

Develop Gen AI agents that will generate new drug product labels or update existing drug product labels for all countries in the world.

**Beta MVP Features (PoT):**

* Generate PILs from SmPC (and other inputs from above) based on QRD template
* Output to DOCX
* English only
* QC process

**MVP Features (S2 release):**

* Generate SmPC from CDS (and other inputs from Figure 1 above)
* Output to DOCX
* English only
* Downstream markets that use English SmPC and PIL as reference
* PMI from USPI
* Multilingual translation

**2024 to 2025 Feature Release Schedule:**

* Output to HL7 FHIR (XML and JSON)
* Multi-label generation (various outputs)
* Japan: Package Insert (JPI)
* Canada: Product Monograph and Patient Medication Information
* US: United States Prescribing Information (USPI)
* Within-label, cross-label and external-to-label Advanced Analytics and Big Data

# Test Scenarios

## Test Scenario #1 – Generate a new PIL

**Scenario:**

A new SmPC document (PDF) for Comirnaty was recently created. The team must now create a new PIL document (DOCX) for Comirnaty from its corresponding SmPC document.

**Objectives:**

1. **Content Generation:** Generate the Comirnaty PIL using all relevant text from Comirnaty SmPC as the data source.
2. **Content Mapping:** Ensure all relevant content related to text, images and tables from SmPC is converted (if necessary) and placed under the correct sub-headings in the PIL, as per QRD pre‑defined content transfer rules (Refer to Appendix 1). Describe an extensible and scalable architecture to allow mapping of terms and concepts from one document type to another as future global use cases are developed.
3. **Mandatory Text:** Transfer all mandatory content from the SmPC to the PIL and place it under the correct sub-headings with no omissions or deviations.
4. **Plain Language:** Convert scientific language from the SmPC to the plain language (i.e., Grade 5-8 reading level[[2]](#footnote-3)) text required in the PIL template.
5. **Medical Terms:** Convert medical terms from scientific to plain language using the pre‑defined mapping dictionary (Refer to [MedDRA Patient-Friendly Term List](https://www.meddra.org/patient-friendly-term-list)).
6. **Task Orchestration:** Demonstrate an orchestration approach to enable the business logic of the application and allowing complex prompting approaches (e.g. few shots learning, prompt chaining) and importing of content from various sources (e.g. graph database, document types)
7. **Quality Control:** Provide citations in the PIL from the SmPC to help SMEs address any discrepancies, omissions or errors.
8. **User Experience:** Generate a wireframe of the UX screen

**Supporting Documents:**

QRD SmPC and PIL Template (Refer to [EMA Product Information](https://www.ema.europa.eu/en/documents/template-form/qrd-product-information-annotated-template-english-version-104_en.pdf) website)

SmPC to PIL Mapping (Refer to Appendices)

COMIRNATY 3mcg SmPC v1.0.pdf

**Success Criteria:**

1. **Content Generation:** Content generated in the PIL includes all the relevant information related to text, image and tables. LLM generated wording should be >60% accurate in comparison with the manually generated PIL.
2. **Content Mapping:** All relevant/required information from SmPC has been placed under the correct sections and sub-heading in the PIL where relevant, as per QRD pre-defined content transfer rules. Section mapping between SmPC and PIL is 100%
3. **Mandatory Text:** Mandatory text transfer accuracy is 100%
4. **Plain Language:** Plain language meets health literacy based on Grade Level 5-8 standards.
5. **Medical Terms:** All PIL medical definitions (side effects) have been converted from scientific to plain language utilising a Medical Dictionary as a primary source. Accuracy of dictionary utilisation is 100%
6. **Task Orchestration:** Selection of LLM based on best result for a section, model training is quick with few samples.
7. **Quality Control:** Correct citations from SmPC are displayed for each section of the PIL.
8. **User Experience:** User can intuitively interpret the wireframe with no instructions. User can envision accomplishing the task with no missteps and minimal effort.

## Test Scenario #2 - Update existing PIL (modify, add or subtract content)

**Scenario:**

The existing SmPC document for Xeljanz has been updated with a new indication for Ankylosing spondylitis. Content throughout the SmPC has been updated to incorporate the new indication. Now the corresponding Xeljanz PIL also needs to be updated to match by incorporating the new content related to the indication.

|  |  |
| --- | --- |
| **Existing Drug Product Label Documents** | **Drug Product Label Documents updated with the new indication** |
| Xeljanz SmPC v1.0 | Xeljanz SmPC v2.0 (now includes new indication content) |
| Xeljanz PIL v1.0 | *Gen AI needs to create the first draft (i.e., editable text) of the Xeljanz PIL v1.1 to include new indication content* |

**Objectives:**

1. **Change Generation:** Generate a draft of the updated PIL to incorporate the SmPC changes. Ensure it is visually easy to identify any modifications made to the PIL’s content (e.g., highlight to show added or modified content; strike out to show deleted content).
2. **Change Mapping:** Accurately identify what modifications, additions and subtractions are in the updated SmPC and make the corresponding modifications, additions and subtractions with the relevant sections in the PIL.
3. **Cascading Update:** Accurately implement the changes, modifications, additions and subtractions in the SmPC into the corresponding sections of the PIL without impacting other sections not impacted by the change.
4. **Quality Control:** Verify that the updated PIL reflects all changes, modifications, additions and subtractions made on the SmPC. Provide citations for all changes, modifications, additions and subtractions in the PIL­­ to help SMEs identify any discrepancies or errors.
5. **User Experience:** Generate a wireframe of the UX screen

**Supporting Documents:**

QRD SmPC and PIL Templates (Refer to [EMA Product Information](https://www.ema.europa.eu/en/documents/template-form/qrd-product-information-annotated-template-english-version-104_en.pdf) website)

SmPC to PIL Mapping (Refer to Appendices)

Xeljanz\_IR - 15-Nov-2021 - AS indication - 5mg-10mg FCT SmPC v1.0.pdf (Refer to .zip)

Xeljanz\_IR - 15-Nov-2021 - AS indication - 5mg-10mg FCT SmPC v2.0.pdf (Refer to .zip)

Xeljanz\_IR - 15-Nov-2021 - AS indication - 5mg-10mg FCT PIL v1.0.pdf (Refer to .zip)

**Success Criteria**

1. **Change Identification:** Accurately identify all changes, modifications, additions and subtractions between SmPC v1.0 and SmPC v2.0. Accuracy of identification is 100%.
2. **Change Mapping:** All changes, modifications, additions and subtractions in the SmPC have been accurately mapped to the relevant corresponding sections in the PIL. The accuracy of mapping is 100%.
3. **Cascading Update:** Implementation of the changes, modifications, additions and subtractions in the relevant section of the PIL is up to 80%.
4. **Quality Assurance Checks:** Correct citations highlighting the changes, modifications, additions and subtractions in SmPC are highlights for each change or modification in the PIL.
5. **User Experience:** User can intuitively interpret the wireframe with no instructions. User can envision accomplishing the task with no missteps and minimal effort.

# Digital Section

* **User Interface** – Node/React, Stratus (N.A. for PoT) for Production deployment
* **LLM** – use what are available now: GPT-4, Claude v2 (need to be LLM agnostic)
* **Should** have capability to refine output based on human feedback.
* **Orchestration components** – must run on AWS & Pfizer VoX Platform
* **Vector Database** – AWS OpenSearch
* **Infrastructure support- Specify** the infrastructure requirements for the environment build and configuration.
* **Design –** We need to have the high-level design with enablers or accelerators to be listed
* **Tech SME -** Availability of Tech subject matter experts who can provide guidance and support throughout the project**.**
* **Data Usage -** For POC public data will be used and for MVP we would like to utilize non public data also (IP)
* **Support Model -** Model to be (which team will support)
* **Flexibility** to adopt to Pfizer AI standards

# Success Criteria

* Accuracy of output
* Scalability across all current and emerging labeling use cases
* Capabilities to add new functionality
* User friendly interface
* Performance of the app
* Human-in-the-loop QC and model performance
* TCO
* Vendor support model
* References Input
* Flexibility to adopt to Pfizer standards

# Timelines & Phase gate plan

|  |  |  |
| --- | --- | --- |
| **Description** | **Date (Month/Day)** | **Outcomes** |
| Scenario & Requirements sent | 3/22 | Bakeoff Package |
| Q&A Sessions with Vendors | 3/26-3/27 | Explain business expectation of the requirements. |
| Demo of the PoT | 4/22-4/30 | Vendors to have a live demo of the PoT requirements |
| Final Proposal from Vendors | 5/6 | Response from Vendors |
| Selection of Vendor | 5/10 | Pfizer communication to Vendors |

# Appendices

## Appendix 1 – Sample SmPC to PIL content map

The Quality Review Document (QRD) can be found on the European Medicines Agency’s (EMA) website: <https://www.ema.europa.eu/en/documents/template-form/qrd-product-information-annotated-template-english-version-104_en.pdf>

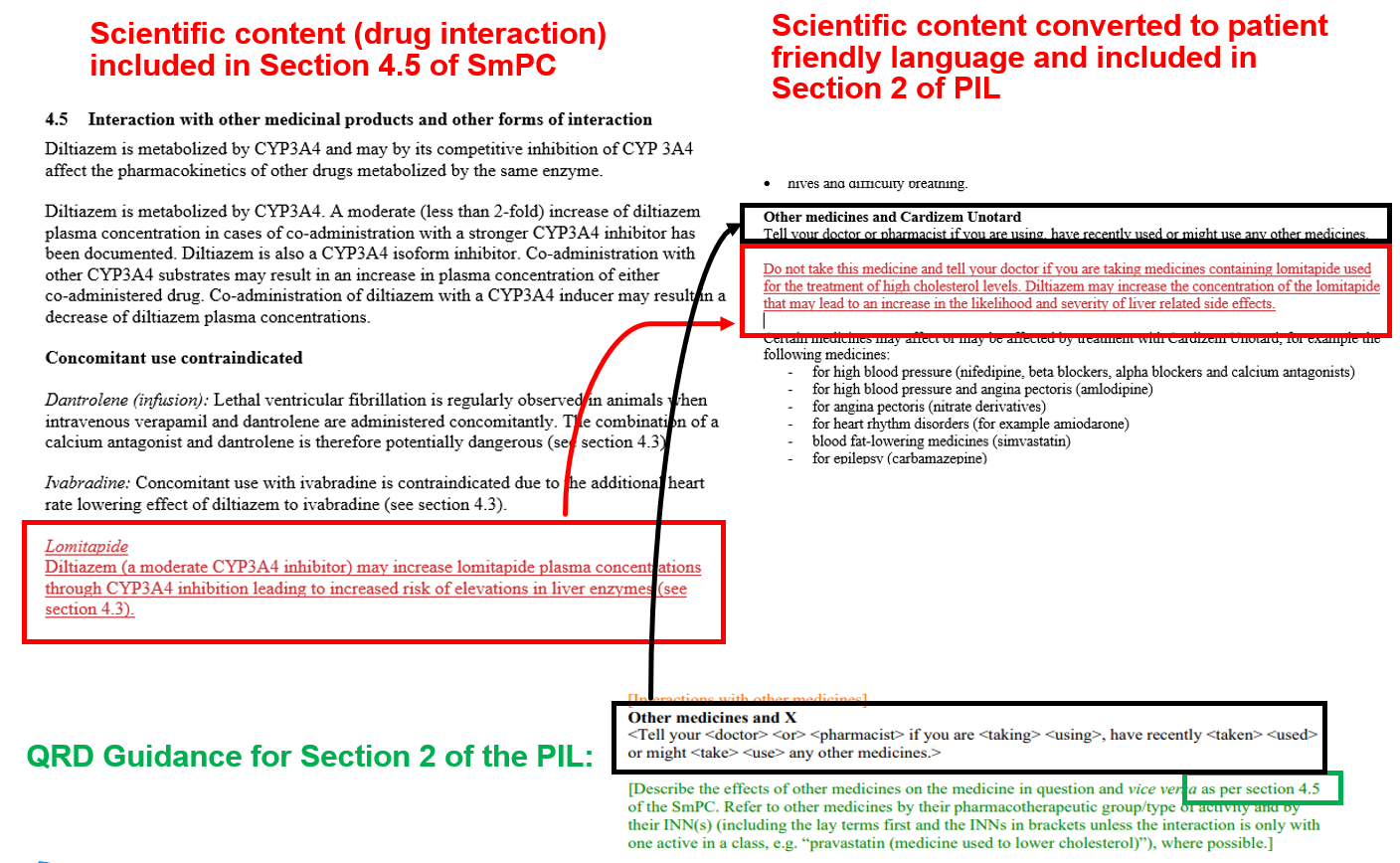
Package leaflet starts at page 25. The PIL starts with Mandatory Text as indicated in QRD (black text). This mandatory Text should be added before the section 1 - "what is in this leaflet".

| **PIL Template Section Headings1** | **PILs source their content from the following sections of the SmPC** | **Mandatory to include this content from the SmPC (See Black text in the QRD template)** |
| --- | --- | --- |
| **1. What X is and what it is used for** |  |  |
| 1. What X is and how it works | 1, 2, 4.1, 5.1 | Yes |
| **2. What you need to know before you take X** |  |  |
| 1. Do not <take> <use> X | 4.3 | Yes |
| 1. Warnings and precautions | 4.4 | Yes |
| 1. Children <and adolescents> | 4.2, 4.4 | Yes |
| 1. Other medicines and X | 4.5 | Yes |
| 1. X with <food> <and> <,> <drink> <and> <alcohol> | 4.2, 4.5 |  |
| 1. Pregnancy <and> <,> breast-feeding <and fertility> | 4.6 | Yes |
| 1. Driving and using machines | 4.7 |  |
| 1. X contains {name the excipient(s)} | 2, 4.4, 6.1 |  |
| **3. How to take X** |  |  |
| 1. Use in children <and adolescents> | 4.2 | Yes |
| 1. If you <take> <use> more X than you should | 4.9 |  |
| 1. If you forget to <take> <use> X | 4.2 | Yes |
| 1. If you stop <taking> <using> X | 4.2, 4.4 | Yes |
| **4. Possible side effects** |  |  |
| 1. Additional side effects in children <and adolescents> | 4.8 | Yes |
| 1. Reporting of side effects | 4.8 | Yes |
| **5. How to store X** |  |  |
| 1. Expiry Date | 6.3 |  |
| 1. Storage Conditions | 6.4 |  |
| 1. Special instructions | 6.3, 6.4 | Yes |
| **6. Contents of the pack and other information** |  |  |
| 1. What X contains | 2, 6.1 | Yes |
| 1. What X looks like and contents of the pack | 3, 6.5 |  |
| 1. Marketing Authorisation Holder and Manufacturer | 7 | Yes |
| 1. This medicine is authorised in the Member States of the European Economic Area <and in the United Kingdom (Northern Ireland)> under the following names: | 7 |  |
| 1. This leaflet was last revised in <{MM/YYYY}><{month YYYY}>. | 9, 10 | Yes |
| 1. Other sources of information | N/a - boiler plate text common to all PILs. Not relevant for this exercise |  |
| 1. The following information is intended for healthcare professionals only: | N/a – not relevant for this exercise |  |

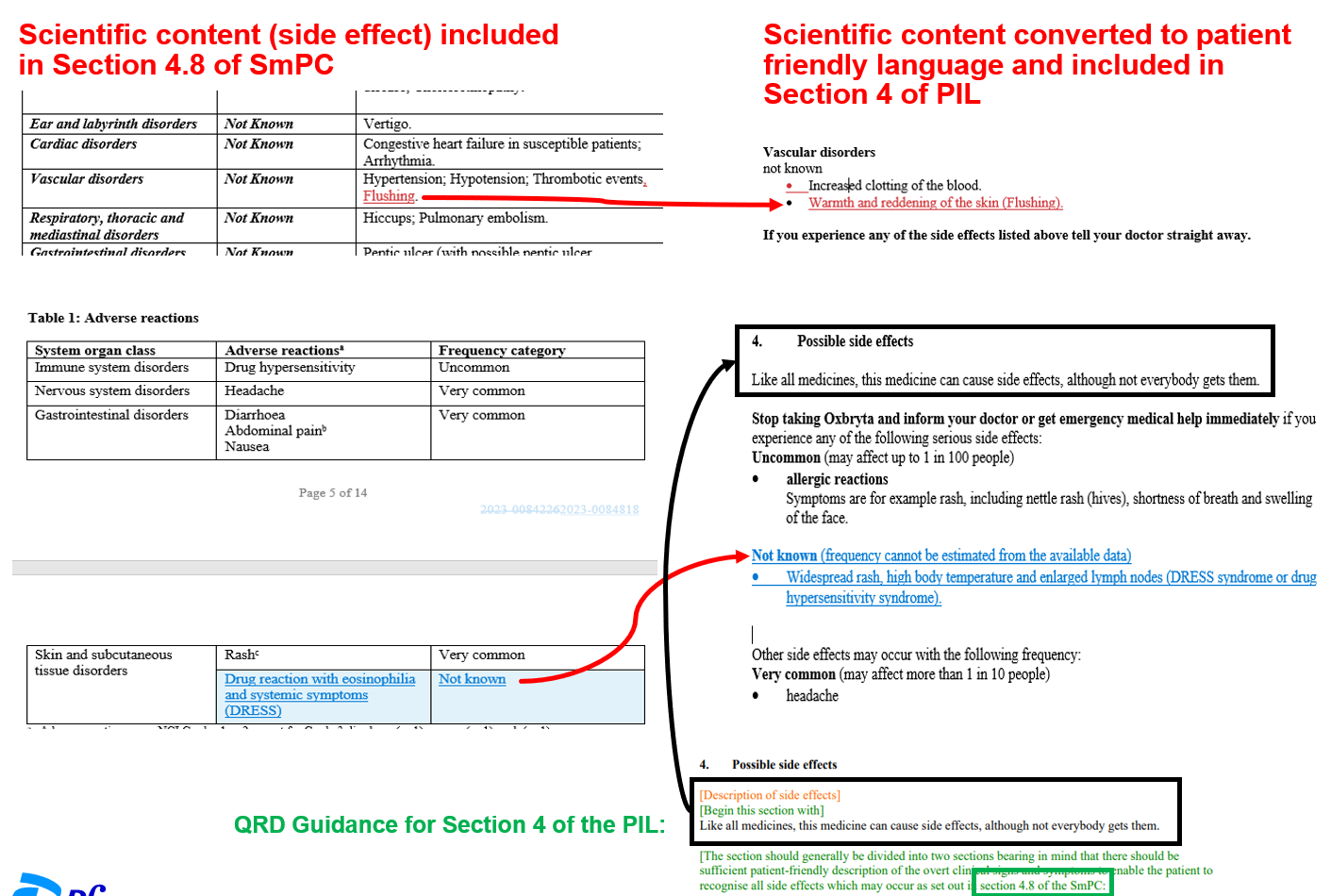
1 “X” is used as a placeholder in the QRD template. “X” is replaced with the name of the drug

## Appendix 2 – SmPC to PIL conversion examples

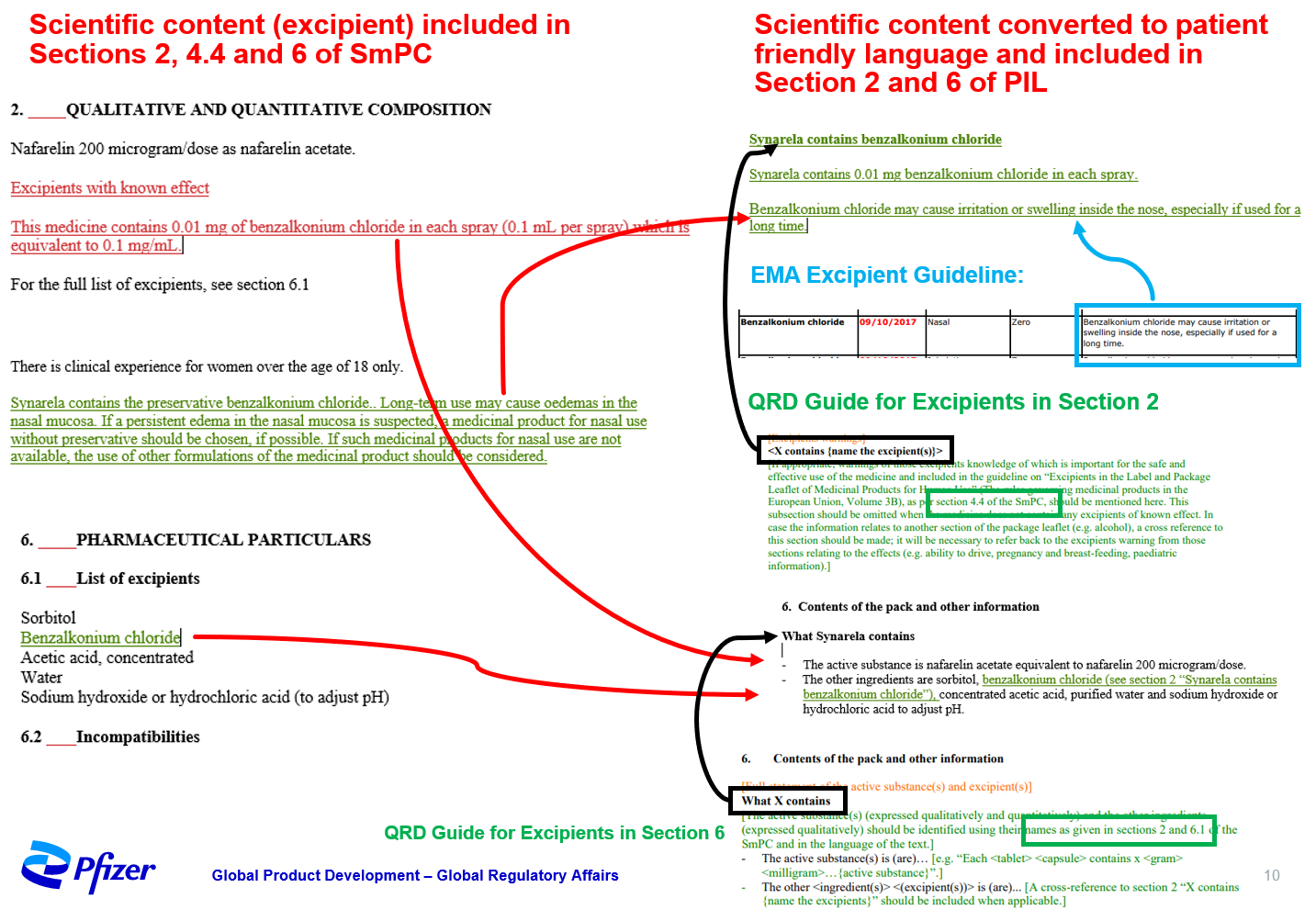
**Figure 2 SmPC section 4.5 converted to PIL section 2**



**Figure 3 SmPC section 4.8 converted to PIL section 4**



**Figure 4 SmPC sections 2, 4.4, and 6 converted to PIL section 2 and 6**



1. Refer to the EMA’s Product Information templates website for further details: [Product-information templates](https://www.ema.europa.eu/en/human-regulatory-overview/marketing-authorisation/product-information-requirements/product-information-templates-human) [↑](#footnote-ref-2)
2. Refer to [Flesch–Kincaid readability tests - Wikipedia](https://en.wikipedia.org/wiki/Flesch%E2%80%93Kincaid_readability_tests) as a guide for readability testing. [↑](#footnote-ref-3)