Vincent Haverhoek
Data Scientist

20 may 2025

How can we utilize the large and rich documents that are the EPAR?

Universiteit Utrecht
Pharmacoepidemiology & Clinical Pharmacology

EPAR uncertainties

Aim of the study:

- What makes an uncertainty section?
- What is said in uncertainty sections?

- Can we harmonize text used in the sections?

3.3. Uncertainties and limitations about favourable effects

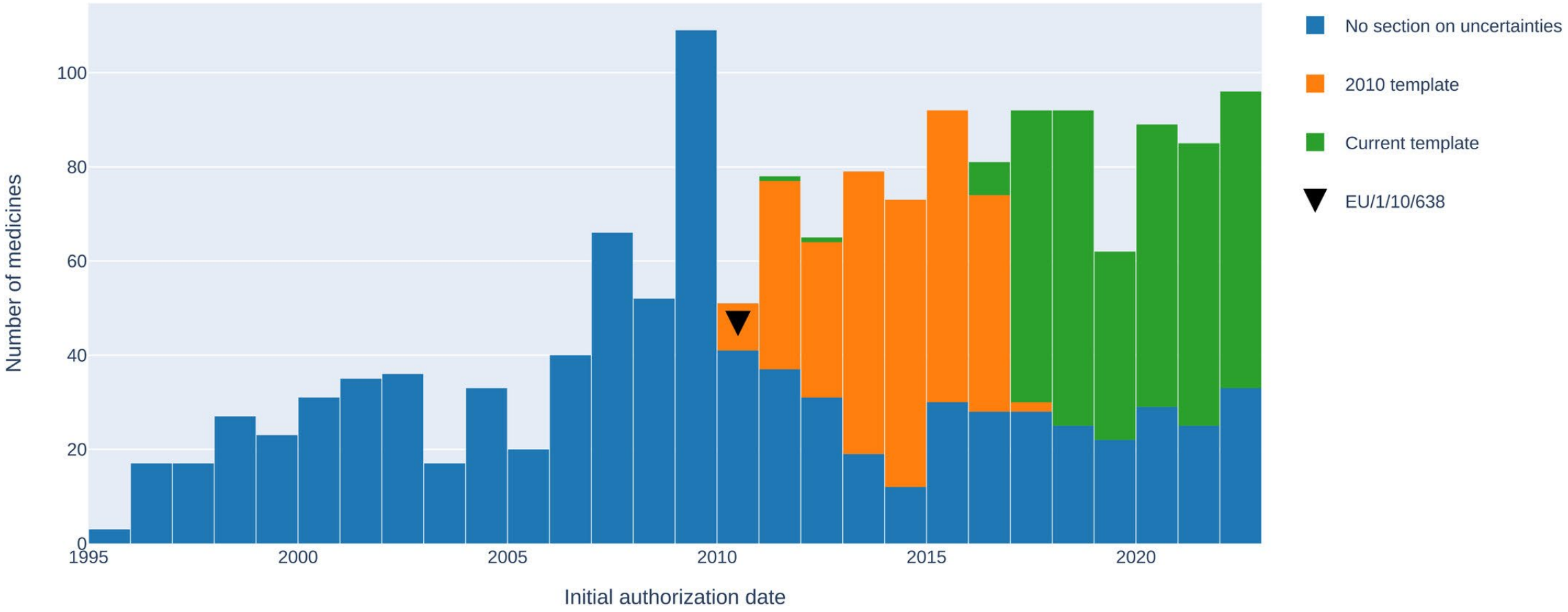
Duration of protection beyond 8 weeks is not known. Efficacy data are not available after this timepoint. Long-term vaccine efficacy data will become available from post-authorisation effectiveness studies and from the ongoing clinical trials. However, participants in the placebo arm are being unblinded and offered vaccination following the FDA EUA for this vaccine. Therefore, it is unclear whether robust efficacy data can be generated. Preliminary immunogenicity results from the FIH trial

3.5. Uncertainties and limitations about unfavourable effects

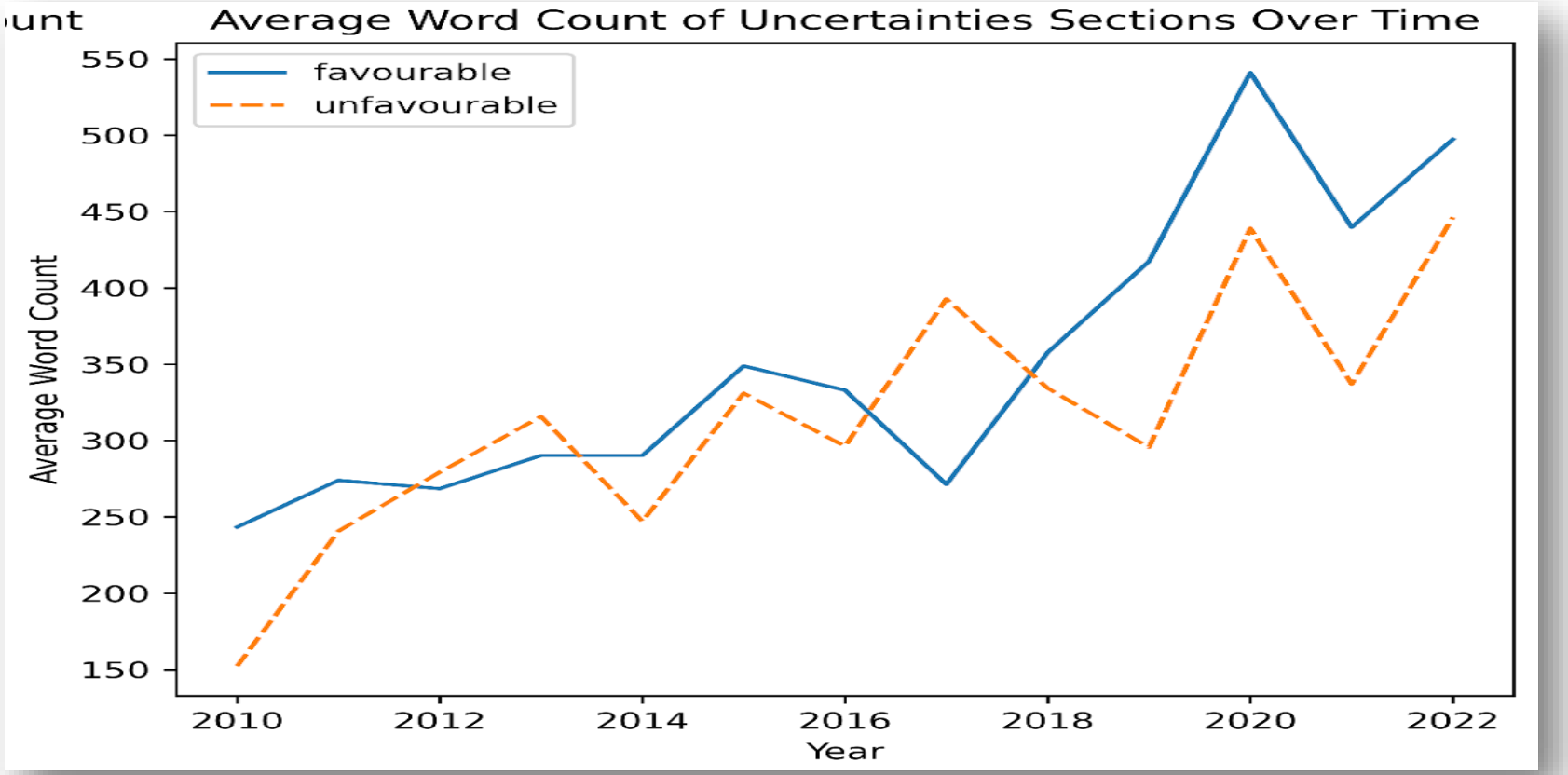
At the time of the primary analysis, the median follow-up after vaccination was 58 days in both groups. Longer safety follow-up of >2 months is available for 23,903 participants in the FAS: 11,948 participants in the Ad26.COV2.S group (54.6%) and 11,955 in the placebo group (54.6%). However, only 34 participants in the Ad26.COV2.S group (0.2%) and 31 in the placebo group (0.1%) have been followed for up to 4 months. Long-term safety data is not yet available and will be characterised as

Jcovden (previously COVID-19 Vaccine Janssen)

EPAR – Where are the uncertainty sections?

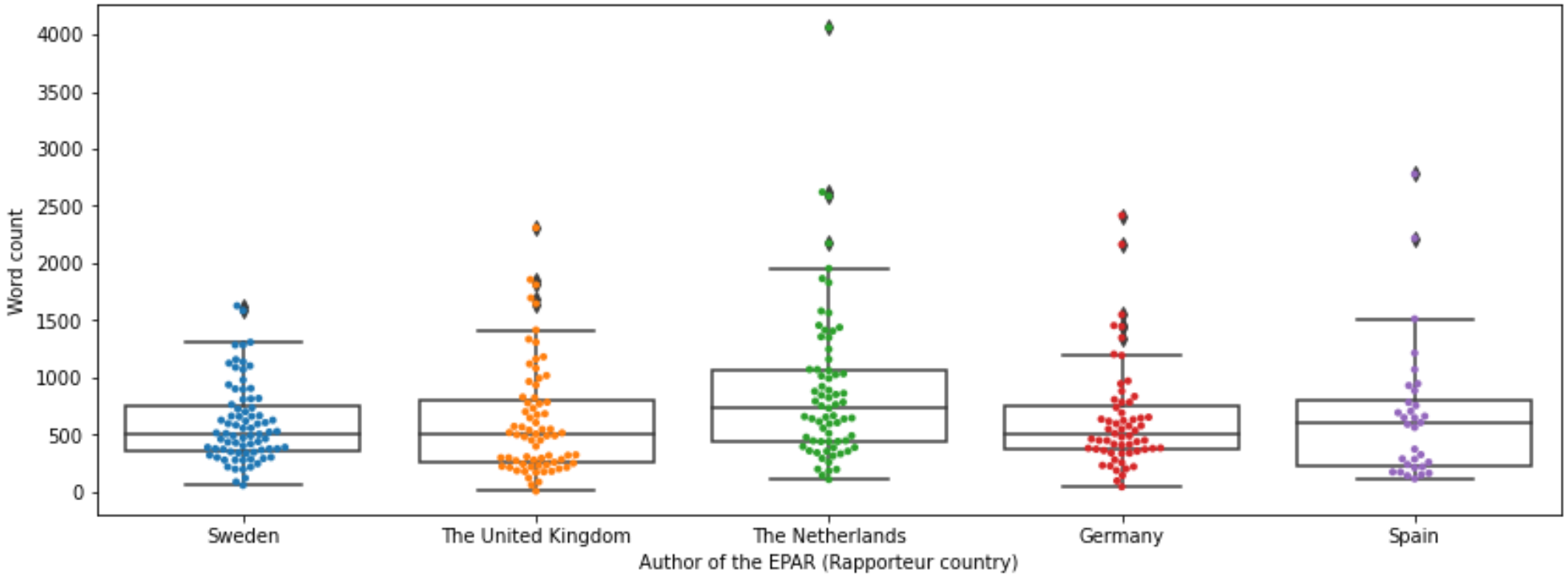


EPAR – How long are the sections?



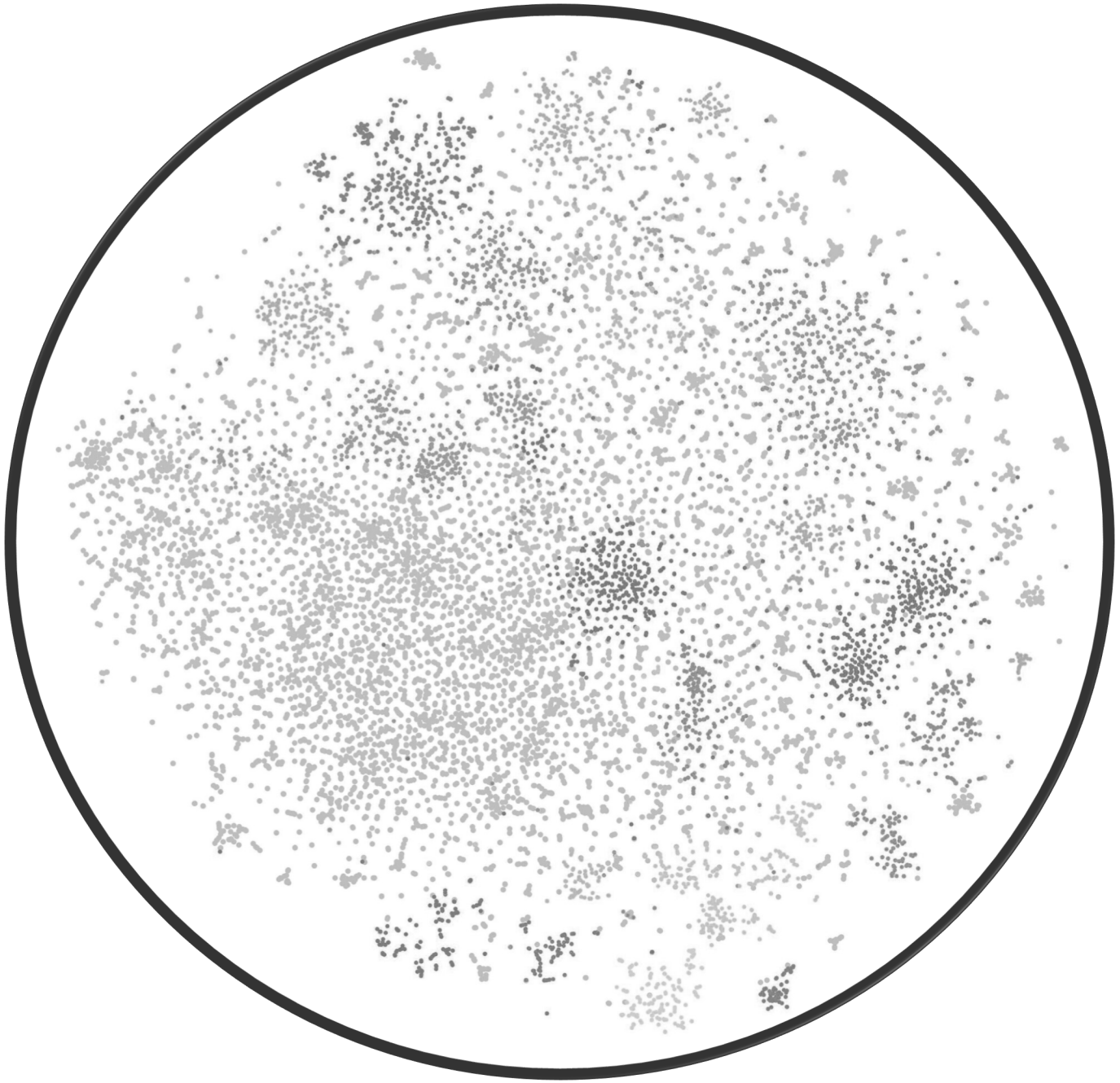
EPAR – How long are the sections?

based on rapporteur country



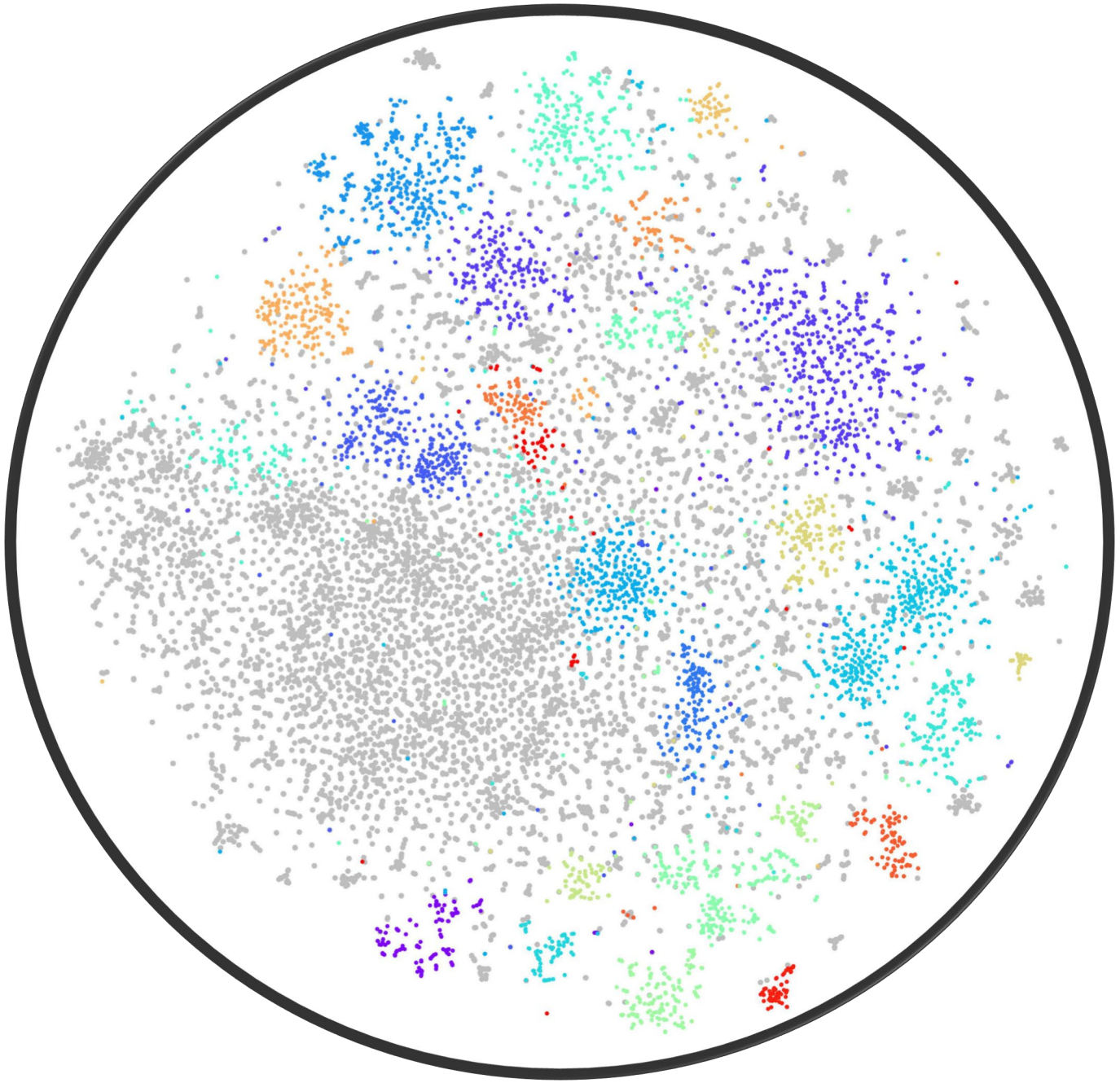
AI analysis

Each dot is a sentence
(+- 15.000 sentences)



AI analysis

Each dot is a sentence
(+- 15.000 sentences)



Example group:
Reference to SmPC

(40+ unique ways to reference)

```
Sentences:
-this is stated in the smpc.
-this is stated in the smpc.
-this is addressed in the rmp
-this was included in the rmp.
-this is addressed in the rmp.
-this is reflected in the smpc.
-this is reflected in the smpc.
-this is reflected in the smpc.
-this is reflected in the smpc.
-this is highlighted in the smpc.
-this is now reflected in the smpc
-this is now reflected in the smpc
-the above is reflected in the smpc.
-this has been reflected in the smpc.
-this should be mentioned in the smpc.
-a warning has been included in the smpc
-these issues are mentioned in the smpc.
-this limitation is reflected in the smpc
-this is adequately reflected in the smpc.
-this is adequately reflected in the smpc.
-this information is reflected in the smpc.
-this information is reflected in the smpc.
-this is sufficiently addressed in the smpc
-adequate measures are included in the smpc.
-this is reflected in section 5.1 of the smpc
-this is included in section 4.8 of the smpc.
-appropriate guidance is included in the smpc.
-this has been adequately reflected in the smpc
```

Example group:
Age

Table 3. Selection of sentences describing uncertainties from three exemplary clusters[16-29]

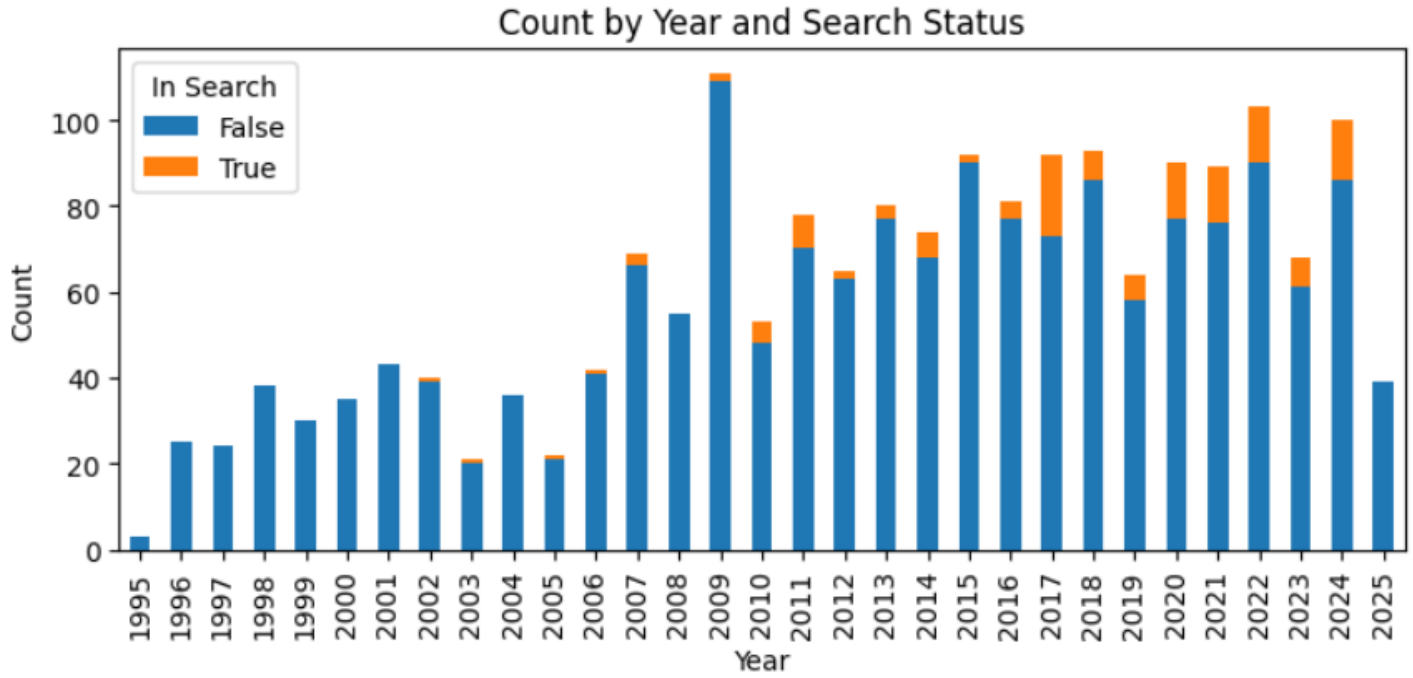
Cluster ID / INN	*N* sentences	*N* unique innovative medicines	*N* unique active substances
Uncertainties related to age			
4	389	185	180
Sotrovimab	"Due to the small sample size of participants > **85 years of age**, no meaningful clinical conclusion can be drawn for that population."		
Defatted powder of *Arachis hypogaea* L., semen (peanuts)	"However, the numbers (especially of **12–17-year-old subjects**) are small and a distinction in efficacy between **age groups** cannot be made."		
Ozanimod	"No (controlled) safety data are available for **pediatric** subjects (< **18 years of age**) and **elderly** subjects (> **55 years of age**)."		
Lenvatinib	"There are no data on the use of lenvatinib in **pediatric population**."		
Pandemic influenza vaccine (H5N1) (split virion, inactivated, adjuvanted)	"However, it is not possible to predict whether half the **adult** dose would suffice in **children aged 3 years**."		
Uncertainties related to pharmacovigilance and lack of safety data			
11	165	130	129

Topic-related keywords within the sentences are shown in bold.

INN, international nonproprietary name.

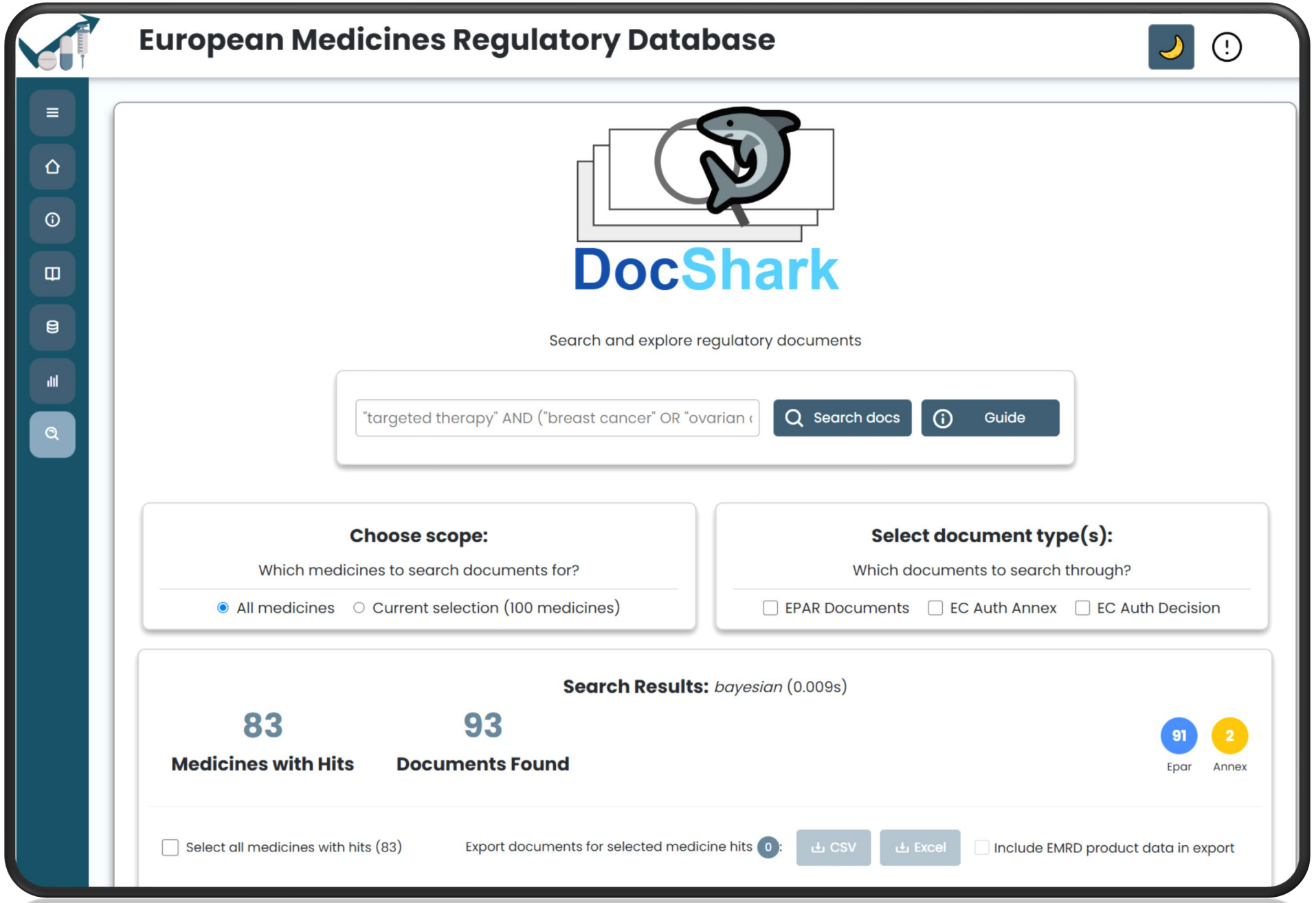
24 relevant groups

✨ Bonus ✨ EPAR use case:
Finding a study cohort using EPARs



Finding a Cohort

- Usage of AI models
- Drug repurposing (C.I.6.a)
- Usage of advisory groups
- Lipid Nano Particles (LNP)
- Progressive multifocal leukoencephalopathy (PML)



Utrecht University

CBG MEB
COLLEGE TER BEOORDELING VAN GENEESMIDDELEN

EMRD

Release 2025

Conclusion - 🧪 EPARs as a Data Source

- **Basic AI/NLP methods** make large-scale analysis feasible and scalable.
- **Significant variation** across reports — domain knowledge is essential for interpretation.
- **High potential** to support regulatory and scientific research.