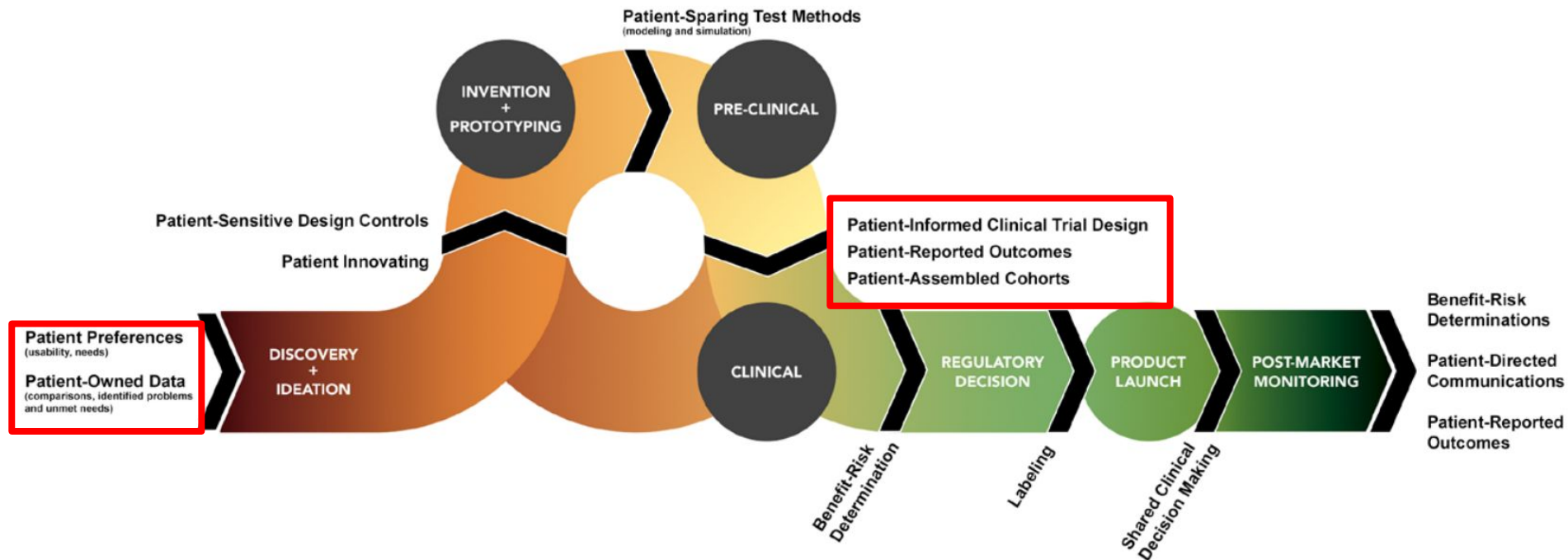


Social Media Listening: A different Approach to Patient Insights

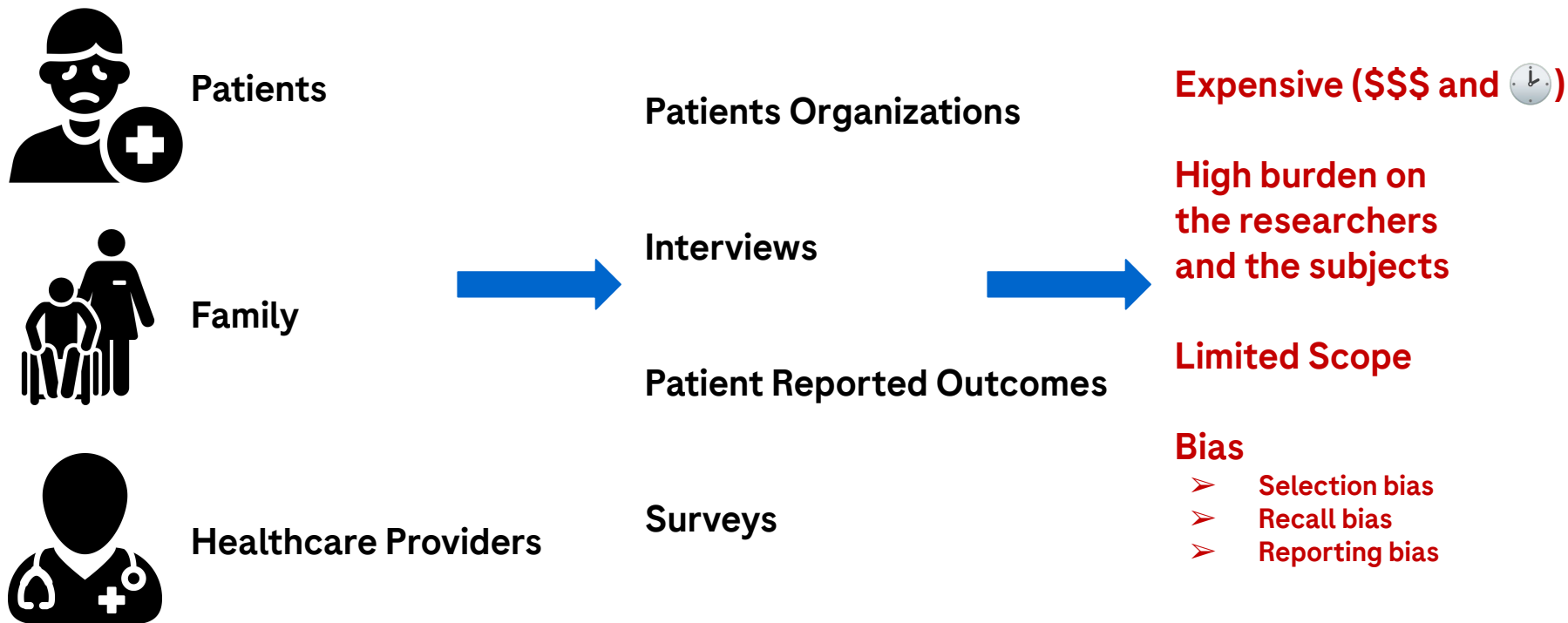
The Turing-Roche Knowledge Share Series

Lucia Schmidt, PhD
PostDoc, Data & Analytics
Pharmaceutical Research & Early Development
lucia.schmidt@roche.com

Patient Centric Drug Development



Traditional Approaches to Patient Insights



Where Social Media Listening Shines



Social Media Listening: passively monitoring users on online

- No Recruitment

Expensive (\$\$\$\$ and 🕒)

- Larger Cohorts

Expected work on less researchers
and the subjects

- (Almost) Real time data

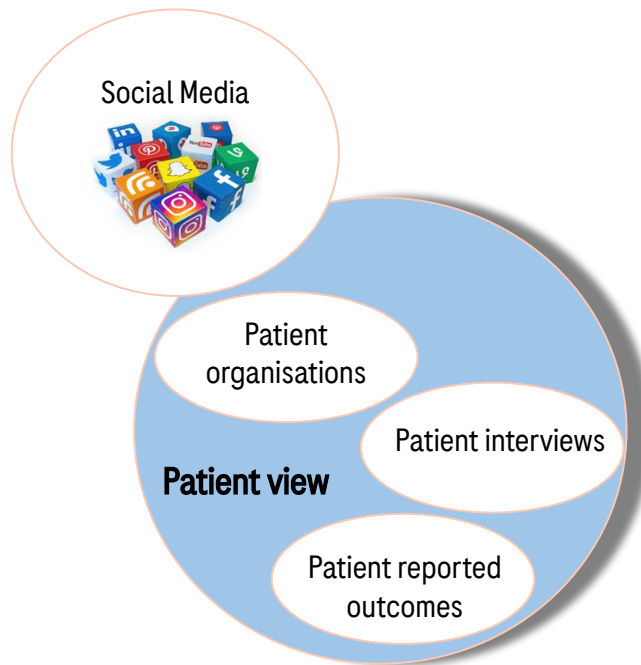
Limited Scope

- Insights come straight from the patient's mouth

Bias

- Selection bias
- Recall bias
- Reporting bias

Where Social Media Listening Stands



Schmidt, Rodriguez-Esteban, Gottowik, Leddin.
Drug Discov Today. 2022.

	Traditional methods	Social media listening
Type of data collected	Mainly qualitative but quantitative in some cases such as group concept mapping	Qualitative
Number of patients	Low	Medium to high
Data time range	Limited to the study	Flexible
Geographical constraints	Yes (unless done online)	None, except for data protection restrictions
Interaction with the patient	Method dependent	No
Ability to explore areas of interest to the researcher	High. If there is patient interaction this facilitates discussion and probes into areas of interest	Medium. Based on the data available
Ability to discover areas of interest to the patient	Medium. Only if the process allows unbiased input from the patient	High. Data exploration allows the discovery of key patient concerns
Availability of clinical or background data	High	Low
Ability to confirm diagnosis	High	Medium, self-confirmed only
Ability to confirm demographic	High	Medium, self-reported only
Level of burden on the subjects	High	Low
Level of burden on the recruiters	Variable. Depending on the method	Low
Level of burden on the researchers	High	High
Time and cost burden	High	Low
Difficulty in extracting information from data	Low to medium	High when involving large volumes
Cost of following up with the patients for future studies	High	N/A
Biases at play	Selection bias, reporting bias, recall bias, framing bias	Selection bias (in terms of social media presence), negativity bias
Scientific acceptance	High	Low
Regulatory support	High	Mixed. Depending on the objective

The Rise of Patient Centricity



New clinical trial regulation framework effective since Feb 2022

Requirement to document layperson and patient involvement in clinical trial design for every newly registered trial



Medicines & Healthcare
products
Regulatory Agency

MHRA pilots patient involvement in new applications since April 2021

Requests evidence of patient involvement activities for drug development



Patient focused drug development guidelines, all 4 fully effective since Feb 2022

Recommends social media as data source to obtain qualitative AND quantitative data for drug development

Where Social Media Listening Stumbles

- Misspellings
- Layman Language
- Multilingual Content
- Missing Information
- Noisy Data
- Sparse Data
- Activity Imbalance



1 - Where?

Data Collection



Forums



reddit



- Search for **sources**
- Read **terms and conditions**
- Data mining

1 - Where?

Data Collection



The analysis scope is **observational** research only.

Privacy ethics statement



Data was downloaded only from **open health social networking sites** and communities.



No information from restricted data areas has been downloaded.



Downloaded text without user profile data other than users "nick-name".
No aggregation or enrichment of data on an individual has been performed



Only a small number of users within Roche have access to the I2E corpus and the results in a **password protected environment**

Privacy

1 - Where?

2 - Who?

Finding our Patient Population



Application of **inclusion** and/or **exclusion** criteria to extract the relevant cohort

No inference allowed, user must explicitly state the inclusion/exclusion criteria in the text

- Age
- Sex
- Location
- Medication
- Comorbidities
- Disease severity
- Substance abuse
- Unstable medical treatment

1 - Where?

2 - Who?

Finding our Patient Population



Application of **inclusion** and/or **exclusion** criteria to extract the relevant cohort

- Regular expressions
- Ontology mapping
- Machine Learning
 - Named entity recognition
 - Classification
- Hybrid (RegEx + ML)

No inference allowed, user must explicitly state the inclusion/exclusion criteria in the text



Validation of the chosen method is key!

1 - Where?

2 - Who?

3 - What?

Feature Extraction

Use of a variety of NLP methods to extract relevant features

- Targeted
- Exploratory



- Symptoms
- Quality of Life Impacts
- Treatments
- Patient Preferences
- Unmet Medical Needs
- Independence
- Emotional State
- Patient Challenges

1 - Where?

2 - Who?

3 - What?

4 - Why?

Quantify, Analyse and Compare



- ★ **Statistical analysis of the findings**
- ★ **Compare with other data**
 - Literature
 - Insurance Data
 - Other RWD
- ★ **Gain Insights**



Social Media Listening...



Has regulatory agencies support as viable method into patient insight



Partners with traditional methods



Enriches patients insights



Thanks You!
Any Questions?