Supplementary File S1

Table of Contents

Demographic Questionnaire 2
Task Oriented Questionnaire 4
Post-task Questionnaire 16
Interview Questions 22

Demographic Questionnaire

Q1 What is your participation number?

_____

Q2 What is your age?

_____

Q3 What is your occupation?

_____

Q4 What is your highest level of education?

- ○ High School (1)
- ○ College/CEGEP (2)
- ○ Undergraduate (3)
- ○ Graduate Degree (Master's/PhD) (4)

Q5 How would you describe your familiarity with visual interfaces?

- Very Familiar (1)
- Somewhat Familiar (2)
- Slightly Familiar (3)
- Not at All Familiar (4)

Q6 How would you describe your familiarity regarding **adverse drug events (ADEs)** (e.g., what they are, how they occur, and how they are monitored)?

- Very Familiar (1)
- Somewhat Familiar (2)
- Slightly Familiar (3)
- Not at All Familiar (4)

Q7 How would you describe your familiarity regarding **patient reported adverse drug event databases** (e.g., systems such as the U.S. FDA's FAERS or Health Canada's Canada Vigilance database)?

- Very Familiar (1)
- Somewhat Familiar (2)
- Slightly Familiar (3)
- Not at All Familiar (4)

Task Oriented Questionnaire

Q1 What is your participation number?

_____

Q2 Task 1.1

Looking at both datasets, which pharmacovigilance database has a higher overall percentage of missing data?

- ○ More missing data in FAERS (1)
- ○ More missing data in Canada Vigilance (2)
- ○ About equal missing data in both (3)

Q3 Task 1.2

What is the overall percentage of missing data in each database? (Provide the percentage for FAERS and for Canada Vigilance.)

_____

_____

_____

_____

_____

Q4 Post-task 1

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q5 Task 2.1

Within the FAERS dataset, which **reporter type** category (i.e., the type of person or organization submitting the report) has the highest percentage of missing data in the reports they submit? (Identify the reporter category with the most incomplete reports in FAERS.)

_____

Q6 Task 2.2

In FAERS, which **reporter type** category has the lowest percentage of missing data (i.e., which type of reporter's submissions are the most complete)?

_____

Q7 Post-task 2

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q8 Task 3.1

In the Canada Vigilance data, in overall reports, looking into the distribution of missingness based on the seriousness of outcome (Serious vs Non-Serious), which group has the highest missingness?

_____

_____

_____

_____

_____

Q9 Task 3.2

In the Canada Vigilance, name one feature that appears to have the lowest missing rate and one feature that appears to have the highest missing rate in the dataset.

_____

_____

_____

_____

_____

Q10 Post-task 3

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Task 4

Using IRVINE's Comparison View, compare the level of missing data for each of the following attributes in FAERS vs. Canada Vigilance. For each attribute, indicate which database has a higher missing data percentage (or if they are roughly equal).

Q11

4.1 Drug Indication (reason for drug use) – Which system has a higher percentage of missing **Drug Indication** information?

- ○ More missing data in FAERS (1)
- ○ More missing data in Canada Vigilance (2)
- ○ About equal missing data in both (3)

Q12

4.2 Drug Dosage – Which system has a higher percentage of missing **Drug Dosage** information?

- ○ More missing data in FAERS (1)
- ○ More missing data in Canada Vigilance (2)
- ○ About equal missing data in both (3)

Q13

4.3 Treatment Duration – Which system has a higher percentage of missing **Treatment Duration** information?

- ○ More missing data in FAERS (1)
- ○ More missing data in Canada Vigilance (2)
- ○ About equal missing data in both (3)

Q14

4.4 Patient Weight – Which system has a higher percentage of missing Patient Weight data?

- ○ More missing data in FAERS (1)
- ○ More missing data in Canada Vigilance (2)
- ○ About equal missing data in both (3)

Q15 Post-task 4

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q16. Task 5.1

Looking at the FAERS database, which country has contributed the largest number of reports to the dataset??

_____

_____

_____

_____

_____

Q17. Task 5.2

In FAERS, Which country shows the highest overall percentage of missing data in its reports?

_____

_____

_____

_____

_____

Q18 Post-task 5

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q19 Task 6

In 2019 both FDA and Canada Health introduced regulatory measures to improve the quality of adverse drug events reported to the FAERS and Canada Vigilance databases. Taking the period 2019-2024 into mind, identify spikes of increase in missingness in the databases and answer the questions below regarding them.

6.1 What was the percentage of highest missingness spike in FAERS, and which reporter type did contribute the most to this spike ?

_____

_____

_____

_____

_____

Q20 Task 6.2

What was the percentage of highest missingness spike in Canada Vigilance, and which reporter type did contribute the most to this spike ?

_____

_____

_____

_____

_____

Q21 Post-task 6

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q22 Task 7.1

In FAERS database which age group consistently shows high rate of missingness overall in the dataset ?

- ○ 0 -1 Month (1)
- ○ 2 Month – 1 Year (2)
- ○ 3 – 11 Years (3)
- ○ 12 – 17 Years (4)
- ○ 18 – 64 Years (5)
- ○ 65 – 85 Years (6)
- ○ More than 85 Years (7)

Q23 Task 7.2

Again in the FAERS database which gender showed higher missingness overall in the dataset?

- ○ More missing data in male patients (1)
- ○ More missing data in female patients (2)
- ○ About equal missing data in both (3)

Q24 Post-task 7

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q25 Task 8.1

In the FAERS dataset, which **report type** category shows the highest overall missingness across features?

- ○ Spontaneous reports (1)
- ○ Report from study (2)
- ○ Other (3)
- ○ Not available to sender (4)

Q26 Task 8.2

Now switching to the Canada Vigilance dataset, compare the pattern of missingness across report types with the FAERS dataset.

Does the same report type show the highest missingness in both systems, or do the patterns differ? Briefly explain your observation.

_____

_____

_____

_____

_____

Q27 Post-task 8

I found the previous task easy to complete.

- ○ Strongly Agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Post-task Questionnaire

Q1 What is your participation number?

_____

Q2 I was confident about the answers I provided for the given tasks.

- ○ Strongly agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q3 I found it easy to complete all the given tasks.

- ○ Strongly agree (1)
- ○ Agree (2)
- ○ Somewhat agree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat disagree (5)
- ○ Disagree (6)
- ○ Strongly disagree (7)

Q4 I was able to clearly explore and evaluate the **overall completeness** of data and in each pharmacovigilance database using the **main view**.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q5 I was able to clearly explore and evaluate the **distribution of missingness across different categories** (e.g., age, gender, reporter type) using the **detail view**.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q6 I was able to clearly explore and evaluate the **geographic distribution** of reports and levels of missingness across countries using the **Choropleth Map**.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q7 I was able to clearly explore and evaluate the **temporal patterns** of missingness (e.g. spikes, improvements, deterioration) using the **temporal view**.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q8 I was able to clearly compare the patterns of missingness across databases.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q9 I found it easy to connect and interpret information across the various visualizations together.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q10 I found it easy to control the visualizations to see what I wanted to know.

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q11 I found it helpful to have multiple types of information and views integrated into a single interactive visualization tool that I could control..

- Strongly agree (1)
- Agree (2)
- Somewhat agree (3)
- Neither agree nor disagree (4)
- Somewhat disagree (5)
- Disagree (6)
- Strongly disagree (7)

Q13 What did you find worked well within the system?

_____

_____

_____

_____

_____

Q14 What did you find confusing within the system?

_____

_____

_____

_____

_____

Q15 Any additional comments about the visualizations and system.

_____

_____

_____

_____

_____

Q16 Would you like to participate in an interview session? Your answer determines whether you might be asked, not whether you will be asked.

- ○ Yes, I want to be an interview candidate. I realize that I may not be invited. (1)
- ○ Do not contact me at all about the interview session. (2)

Interview Questions

Review of Tasks

[Note: For each participant who engages in an interview, his/her task responses and post-task questionnaire will be examined and certain responses will be chosen as desirable for gathering further information. Those responses will be used below.]

1. How did you go about completing the tasks?

2. On the questionnaire you said <response>. Could you explain this in more detail?

Interview Questions – Experience with IRVINE

[Note: The system will be available for us to use as during the discussion.]

1. After using the IRVINE system, what are your overall impressions of it?
2. Which features of the system did you find most helpful for understanding missing data patterns?
3. Were there any parts of the system or visualizations that you found confusing or difficult to use?
4. How did the tool affect the way you explored and interpreted the data?
5. How well do you think the system supported your ability to connect information across different dimensions (e.g., countries, features, or time)?
6. What improvements or additional features would make the system more useful for you?
7. Do you think IRVINE could be useful for others who work with health data or pharmacovigilance information? Why or why not?

Interview Questions – Final

1. Do you have any additional comments or suggestions?

Thank you for participating in this interview. I am now turning off the recorder.