Best practices for validating literature monitoring searches

Presented by Júlio dos Anjos

November 22, 2017

Agenda

- Literature Monitoring: Why?
- Literature Monitoring: Challenges.
- Using A&I databased to monitor literature
- Building and Validating the Search Formula
  - Scenario with existing Gold set
  - Scenario without Gold set
  - For Investigational New Drugs
- Periodic review
- Caveats
- Q&A time

Literature Monitoring: Why?

Safety information from spontaneous and literature adverse reactions reports may differ

Drug Substance	System Organ Class	Literature Cases (%)	Spontaneous Cases (%)	% Difference
Acetylsalicylic Acid	Nervous System Disorders	25.6	8	17.6
	Gastrointestinal Disorders	8.4	25.4	17.0
Fentanyl	Injury, poisoning, complications	35.9	7.5	28.3
Alendronic acid	Gastrointestinal disorders	4.6	21.0	16.5
	Injury, poisoning, complications	28.3	5.4	22.9
Tamsulosin	Injury, poisoning, complications	50	4.1	45.9
Etoposide	Congenital, familial, genetic	0	24.3	24.3

Klose J, Fröhling S, Kroth E, Dobmeyer T, Nolting A. Safety information from spontaneous and literature adverse reactions reports: a comparison. *Ther Innov Regul Sci.* 2013;47:248–55.

Safety information from spontaneous and literature adverse reactions reports may differ

Drug Substance	System Organ Class	Literature Cases (%)	Spontaneous Cases (%)	% Difference
Acetylsalicylic Acid	Nervous System Disorders	25.6	8	17.6
	Gastrointestinal Disorders	8.4	25.4	17.0
Fentanyl	Injury, poisoning, complications	35.9	7.5	28.3
Alendronic acid	Gastrointestinal disorders	4.6	21.0	16.5
	Injury, poisoning, complications	28.3	5.4	22.9
Tamsulosin	Injury, poisoning, complications	50	4.1	45.9
Etoposide	Congenital, familial, genetic	0	24.3	24.3

Klose J, Fröhling S, Kroth E, Dobmeyer T, Nolting A. Safety information from spontaneous and literature adverse reactions reports: a comparison. *Ther Innov Regul Sci.* 2013;47:248–55.

Safety information from spontaneous and literature adverse reactions reports may differ

Drug Substance	System Organ Class	Literature Cases (%)	Spontaneous Cases (%)	% Difference
Acetylsalicylic Acid	Nervous System Disorders	25.6	8	17.6
	Gastrointestinal Disorders	8.4	25.4	17.0
Fentanyl	Injury, poisoning, complications	35.9	7.5	28.3
Alendronic acid	Gastrointestinal disorders	4.6	21.0	16.5
	Injury, poisoning, complications	28.3	5.4	22.9
Tamsulosin	Injury, poisoning, complications	50	4.1	45.9
Etoposide	Congenital, familial, genetic	0	24.3	24.3

Klose J, Fröhling S, Kroth E, Dobmeyer T, Nolting A. Safety information from spontaneous and literature adverse reactions reports: a comparison. *Ther Innov Regul Sci.* 2013;47:248–55.

Safety information from spontaneous and literature adverse reactions reports may differ

Drug Substance	System Organ Class	Literature Cases (%)	Spontaneous Cases (%)	% Difference
Acetylsalicylic Acid	Nervous System Disorders	25.6	8	17.6
	Gastrointestinal Disorders	8.4	25.4	17.0
Fentanyl	Injury, poisoning, complications	35.9	7.5	28.3
Alendronic acid	Gastrointestinal disorders	4.6	21.0	16.5
	Injury, poisoning, complications	28.3	5.4	22.9
Tamsulosin	Injury, poisoning, complications	50	4.1	45.9
Etoposide	Congenital, familial, genetic	0	24.3	24.3

Klose J, Fröhling S, Kroth E, Dobmeyer T, Nolting A. Safety information from spontaneous and literature adverse reactions reports: a comparison. *Ther Innov Regul Sci.* 2013;47:248–55.

Guidance by authorities

	EMA	FDA	CIOMS	ICH
Frequency of screening literature	At least weekly	Not specified	Monthly	According to local requirements or at least bi-weekly
Which Literature to screen?	Scientific and medical literature	Scientific literature	Discusses all terminology used by different Regulators	Worldwide literature
Reporting requirements	Serious and non-serious	Serious and unexpected	Expedited reporting being discussed	Day 0 is when Medical safety information is identified
Exclusions	No	Yes	Not applicable	Brand or trade name

Literature Monitoring: Challenges

What are the biggest challenges with Literature Screening?

- Differences in Regulations
- Building an Ideal Search Strategy
- Large Volume of Scientific Literature
- Inspections/Audits
- Monitoring Local Language Journals
- Implementing EMA MLM Results

What are the biggest challenges with Literature Screening?

- Differences in Regulations
- Building an Ideal Search Strategy
- Large Volume of Scientific Literature
- Inspections/Audits
- Monitoring Local Language Journals
- Implementing EMA MLM Results

What are the biggest challenges with Literature Screening?

Differences in Regulations

Building an Ideal Search Strategy

Large Volume of Scientific Literature

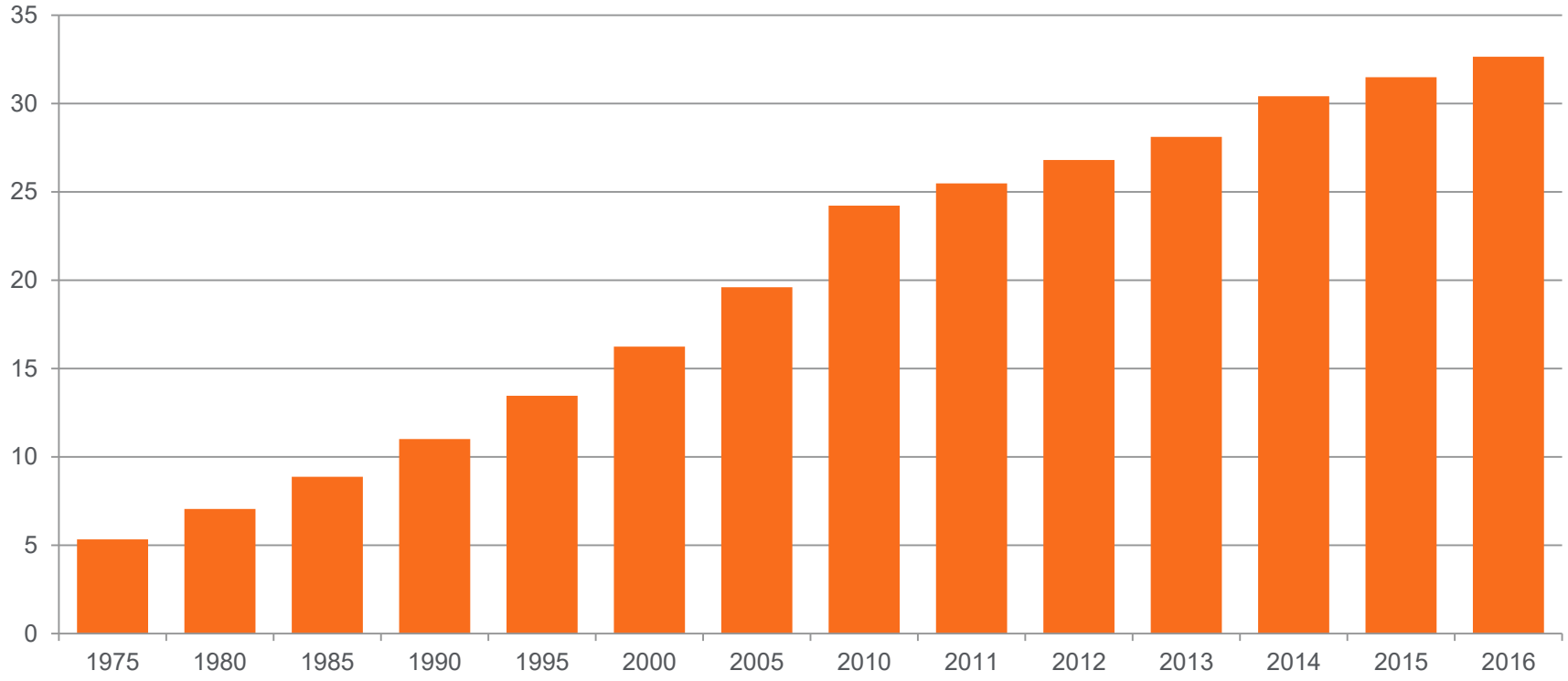
Inspections/Audits

Monitoring Local Language Journals

Implementing EMA MLM Results

Challenges in Literature screening for PV: *Increasing amount of scientific literature*

Number of records in Embase per Year (in **millions**)



Data retrieved from www.Embase.com

Embase®

What are the biggest challenges with Literature Screening?

Differences in Regulations

Building an Ideal Search Strategy

Large Volume of Scientific Literature

Inspections/Audits

Monitoring Local Language Journals

Implementing EMA MLM Results

Using A&I databases

What can you do?

Read all literature – Impossible

Use an A&I database

- Read all PubMed – Impossible
- Read all Embase – Impossible

What can you do?

Remaining option:

- Find and read **only relevant** records from literature (A&I) databases
- Read the articles these records represent

Building the search

Objective:

For any given A&I database that you use:

- Build a search formula that
  - Retrieves all ADR/AE records
  - Retrieves as few as possible non-ADR/AE records
- Validate the search formula

Validate the formula: Why?

- You must be ready to answer: “Why use this formula and not another”, with documented procedures and results…

- “Establishing documented evidence which provides a high degree of assurance that **a specific search formula** will consistently produce **a record set** meeting its pre-determined specifications and quality attributes”

Quality attributes of a search formula:

- **Precision**: How many of the records in the result set are relevant to my need (and, as a consequence, how many are not relevant).

- **Recall**: From all records that are relevant to our need, how many are present in the search result, and how many, if any, are relevant, but the search formula failed to capture.

Quality attributes of a search formula:

- In PV we want a **Recall** of 100%: NO RECORDS indicating that an ADR is present in the article, are missed by the search formula

- **Precision** will impact productivity: Low values mean too much noise is coming in: irrelevant records to review. But it can't be increased at the cost of **Recall.**

Validating the Search Formula

Building and Validating the Search: How

Scenarios:

I: You have previous information about articles that contain AE's: **a Gold set**

II: You have **no Gold set**.

III: It's a new investigational drug, no articles exist

Aspects:

- Formula creation
- Risk Assessment of outliers
- Regular Review of Formula

Scenario I: A Gold set exists

Step 1: Gather a Gold set

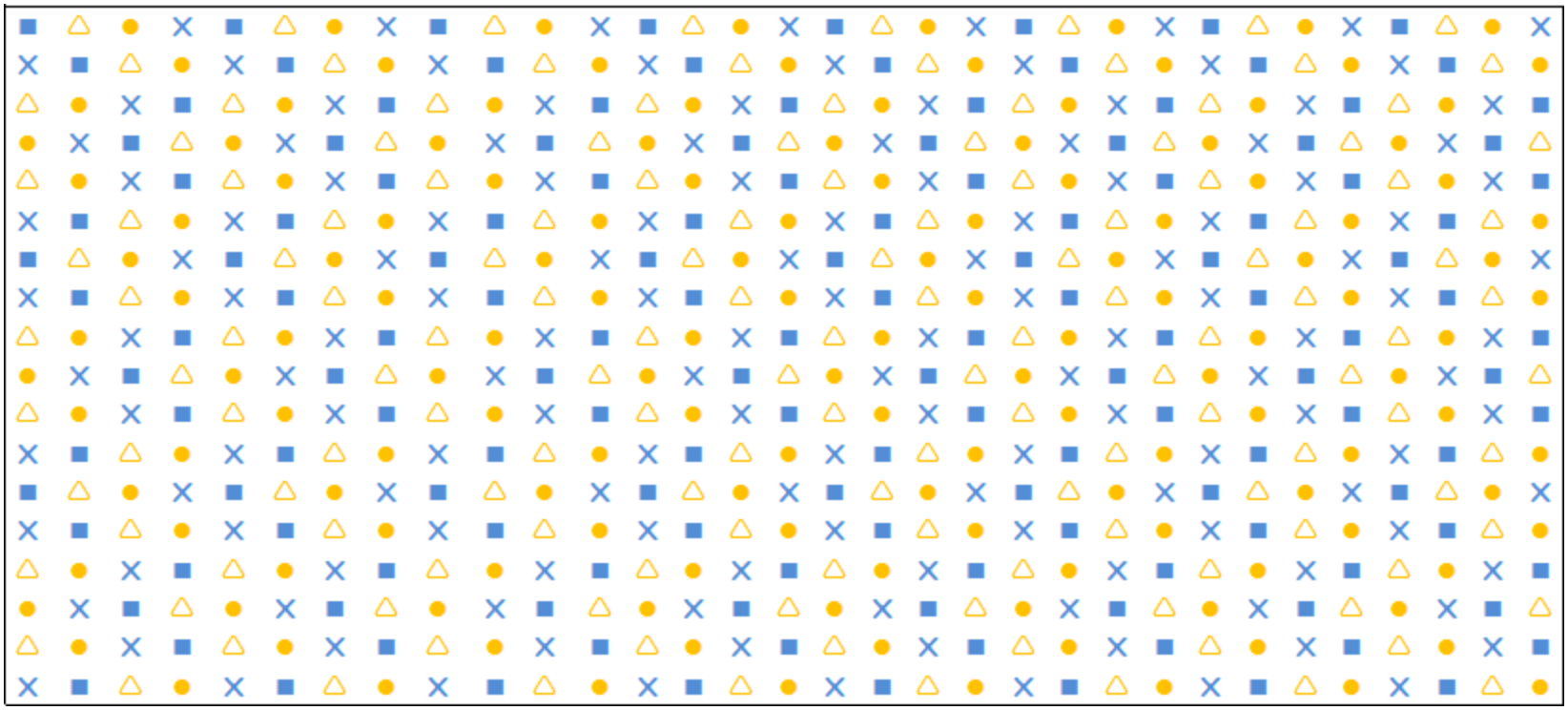
You have previous information about articles that contain ADR/AE's

Create a Gold set from all known relevant articles

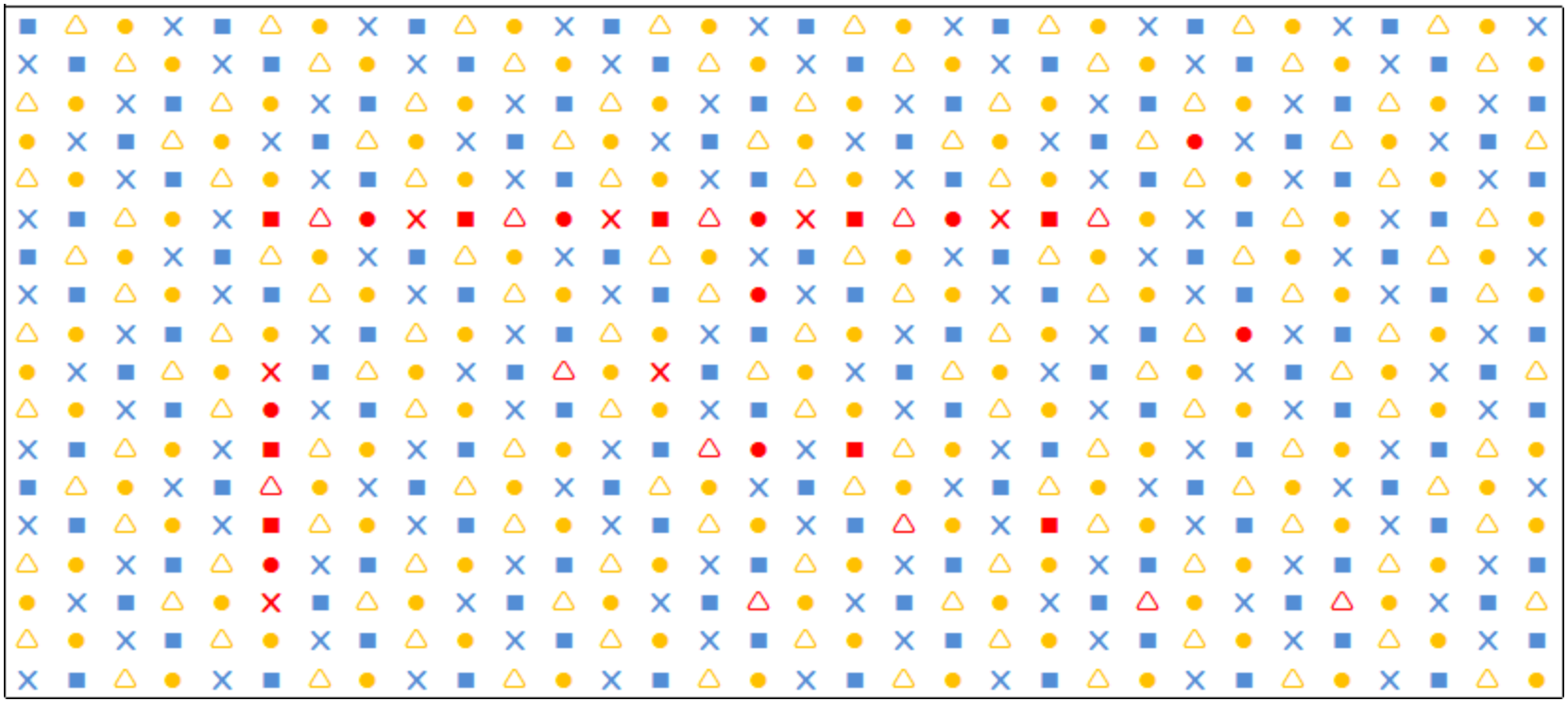
- For a limited timebox. Ex: 2014-2016
- Extract no less than 150 Articles
- Extract no more than SQRT(N)+1
  where N = Total amount of records from 2014 to 2016
- Locate the records for all articles in the A&I database

Graphical representation

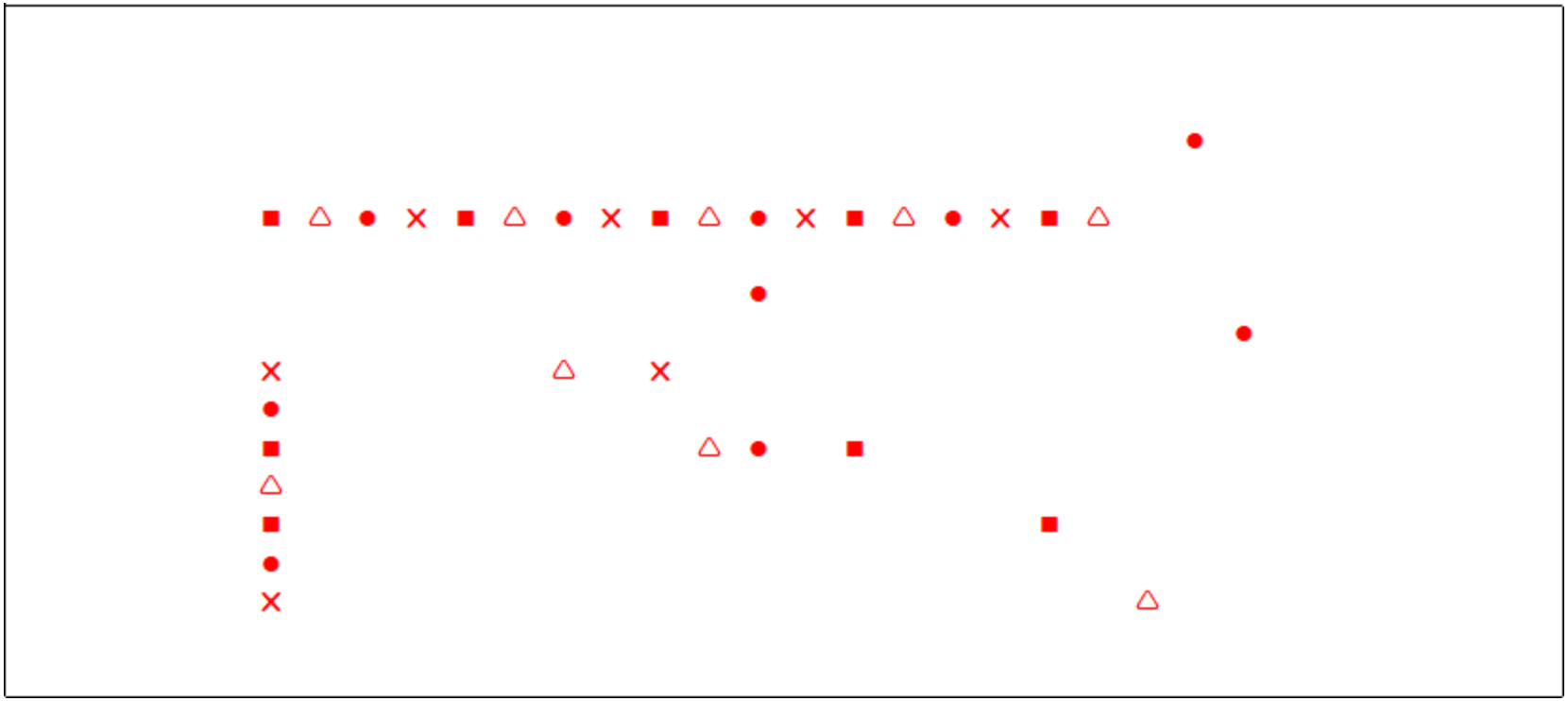
- All database records for 2014-2016



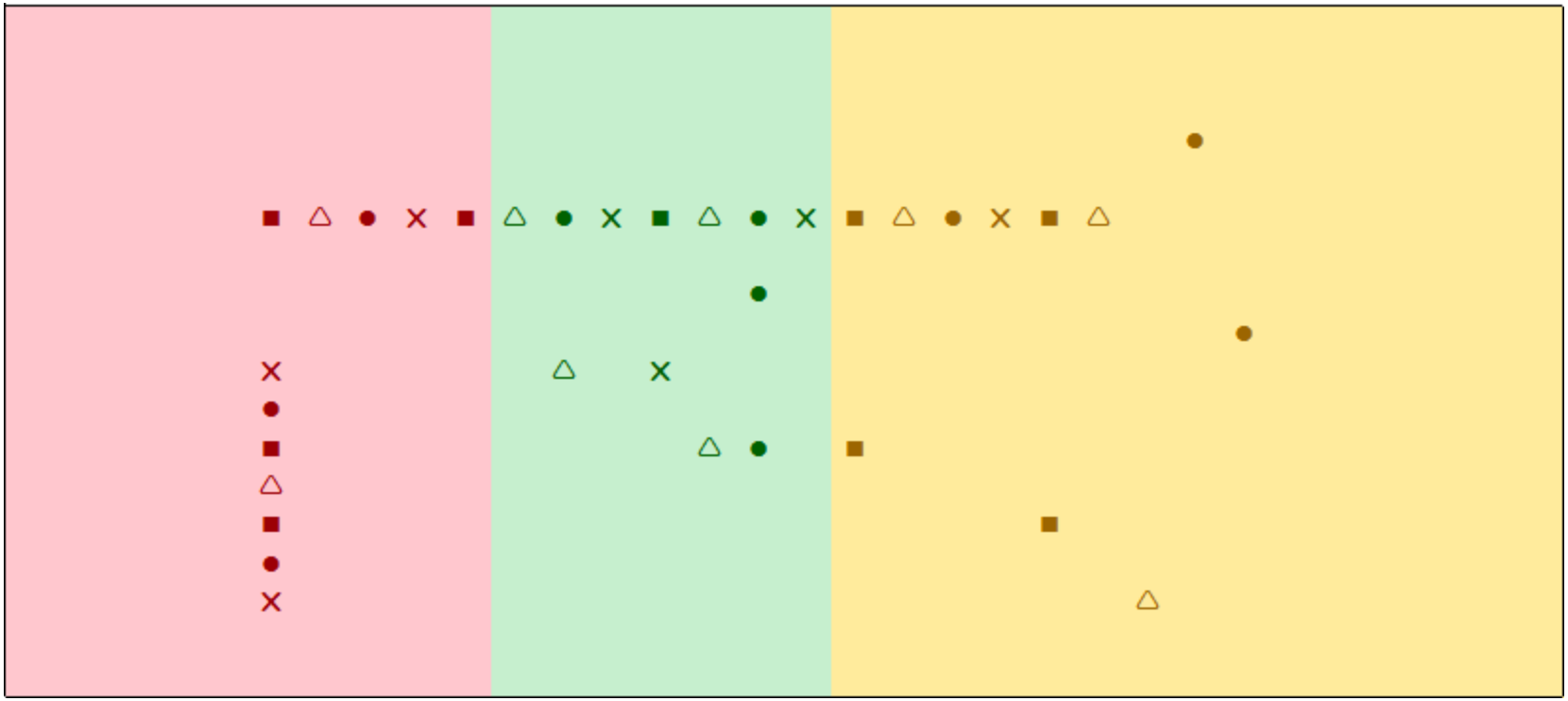
Locate Gold set records



Isolate records for articles in Gold set



Split it in 3 sets: A-B-C



Step 2: Build a formula for set A

- Locate the records for each article in set A.
- Find commonalities in all records.

Step 2: Build a formula for set A

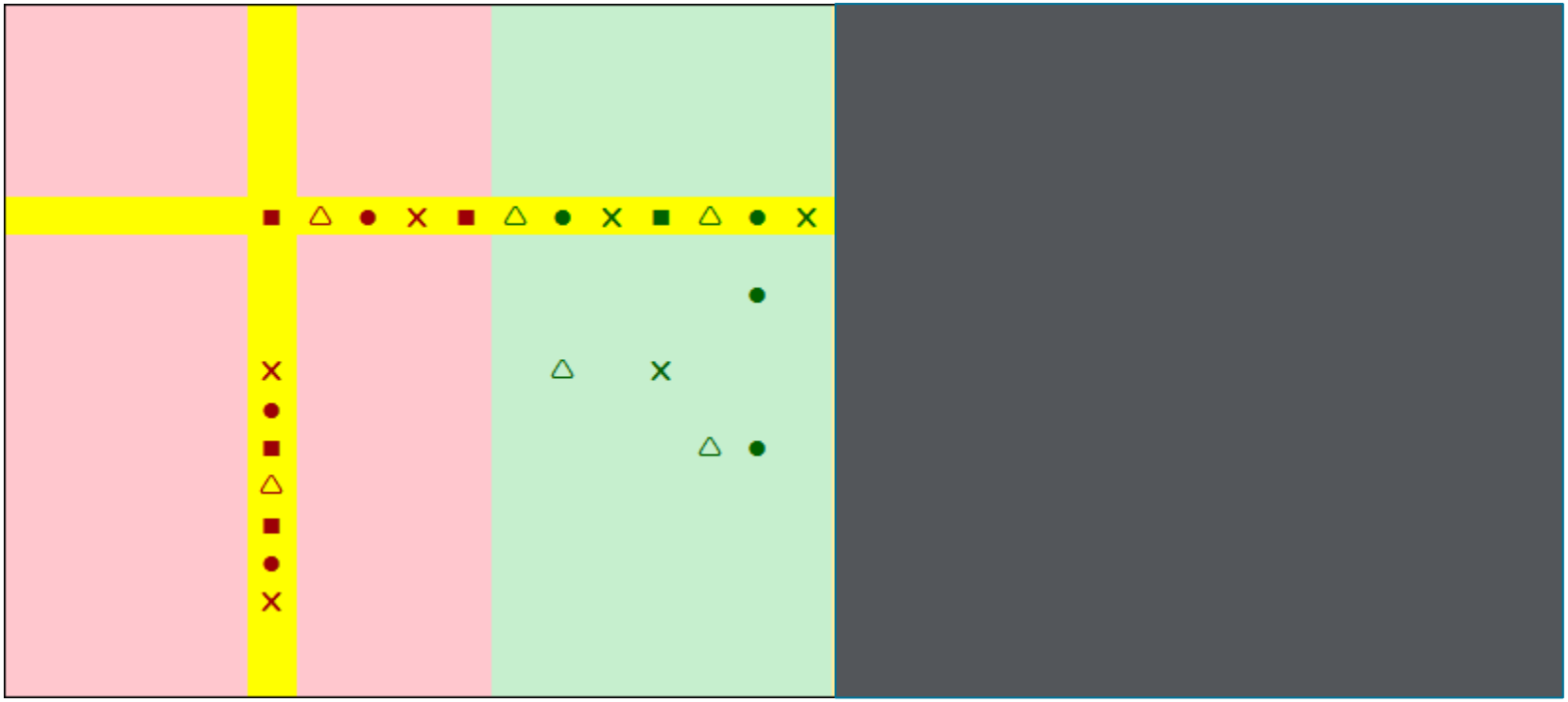
- Build the search formula from this analysis.

Step 2: Build a formula for set A

- Build the search formula from this analysis.

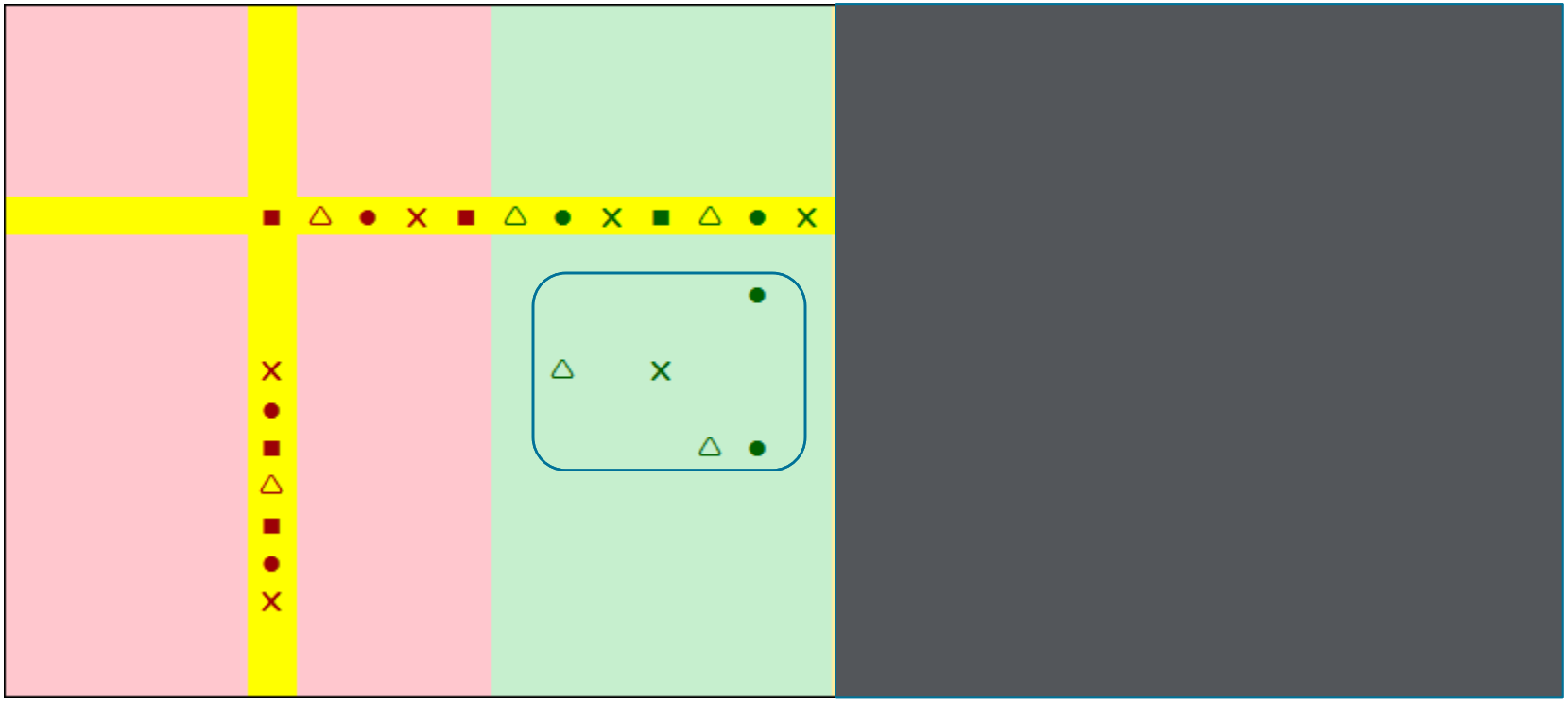
Step 3: Review formula with set B

- Apply search formula to database
- Are all records from set B in result?



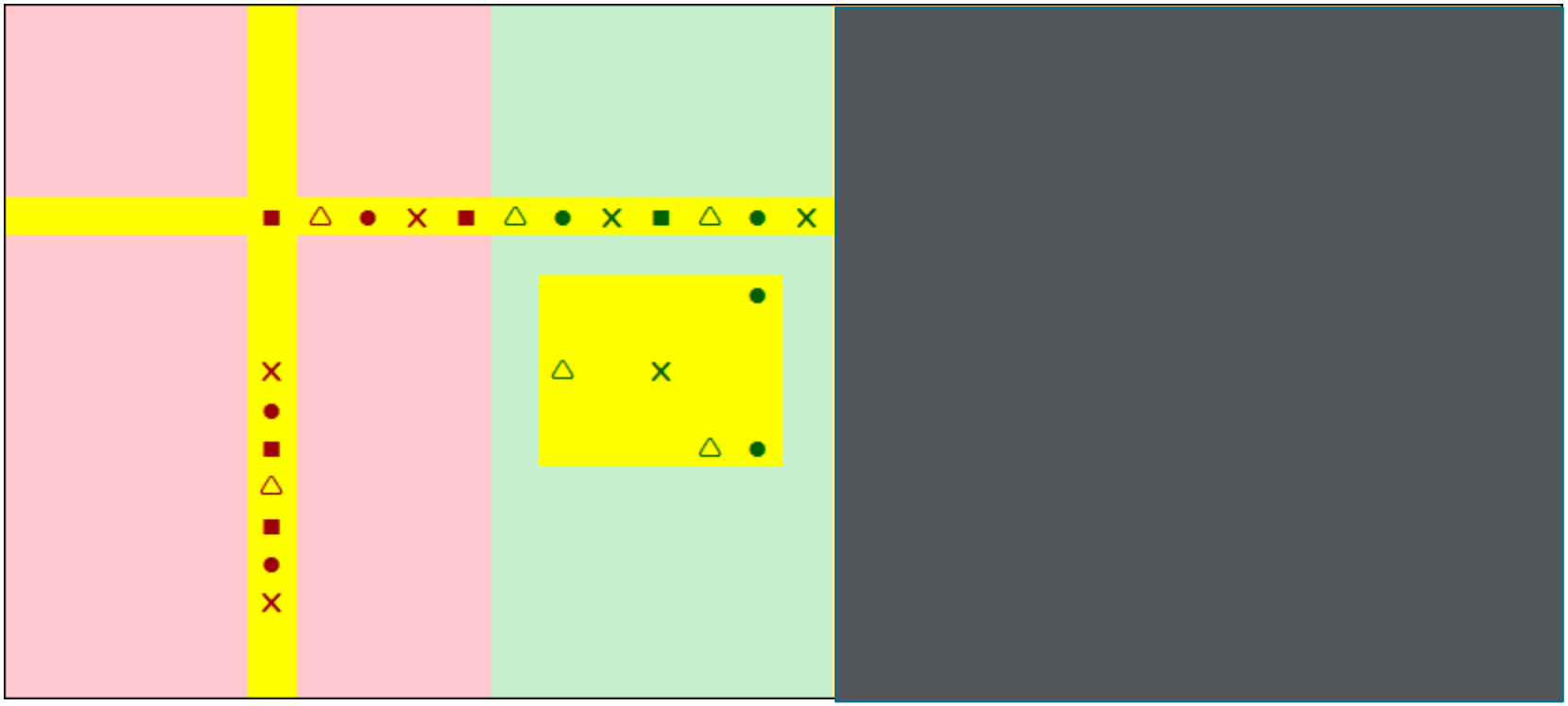
Step 3: Review formula with set B

- Apply search formula to database
- Are all records from set B in result?



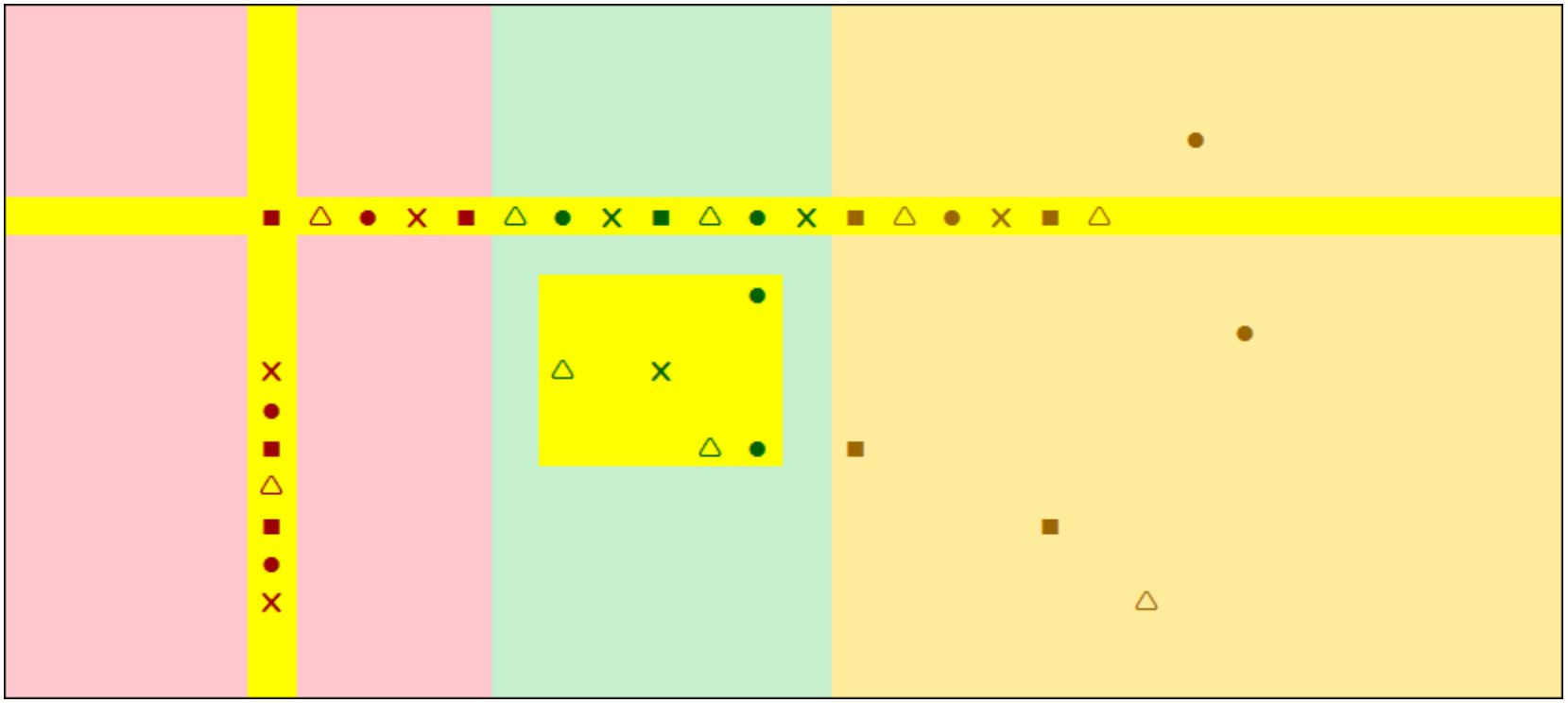
Step 3: Review formula with set B

- Review those not in the result.
- Revise search formula



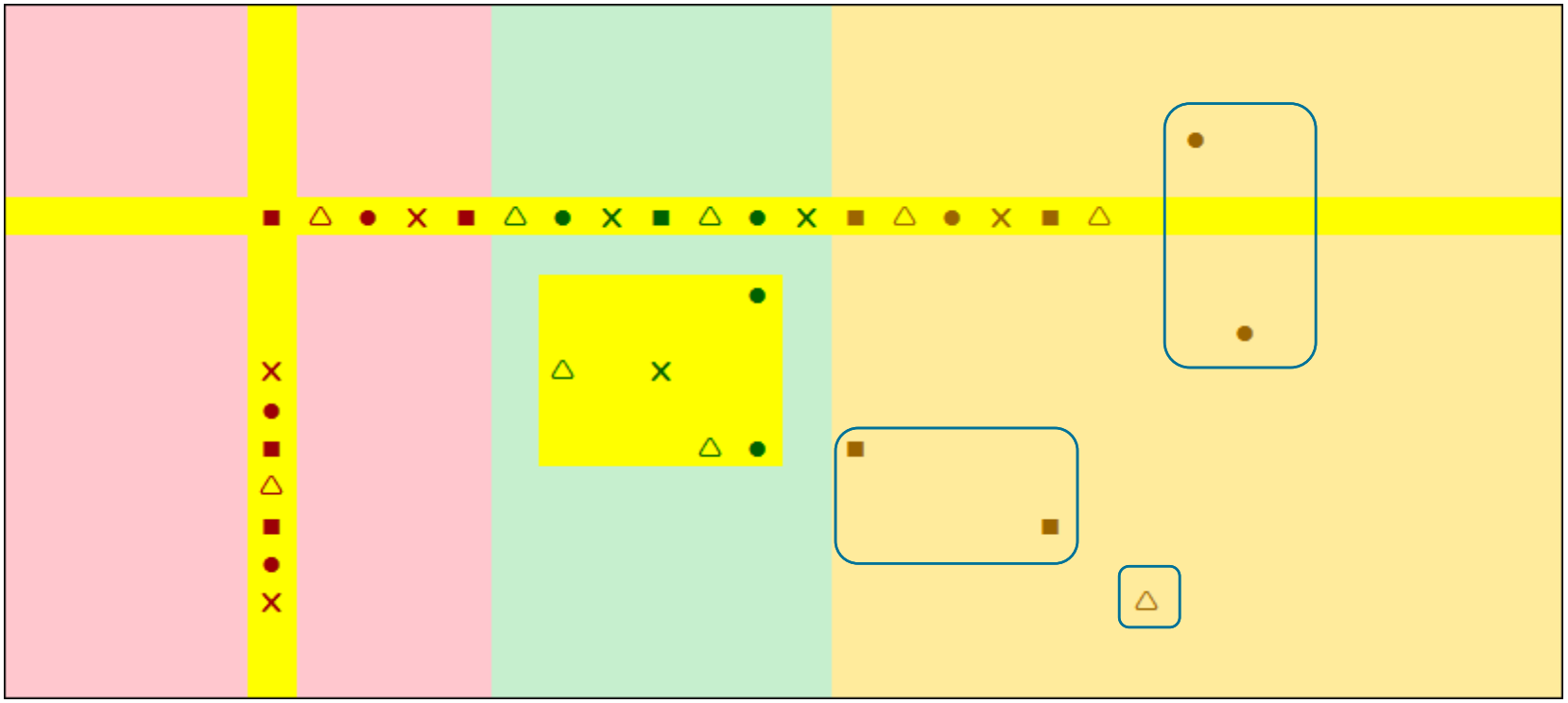
Step 4: Review formula with set C

- Apply search formula to database
- Are all records from set A and B in result?



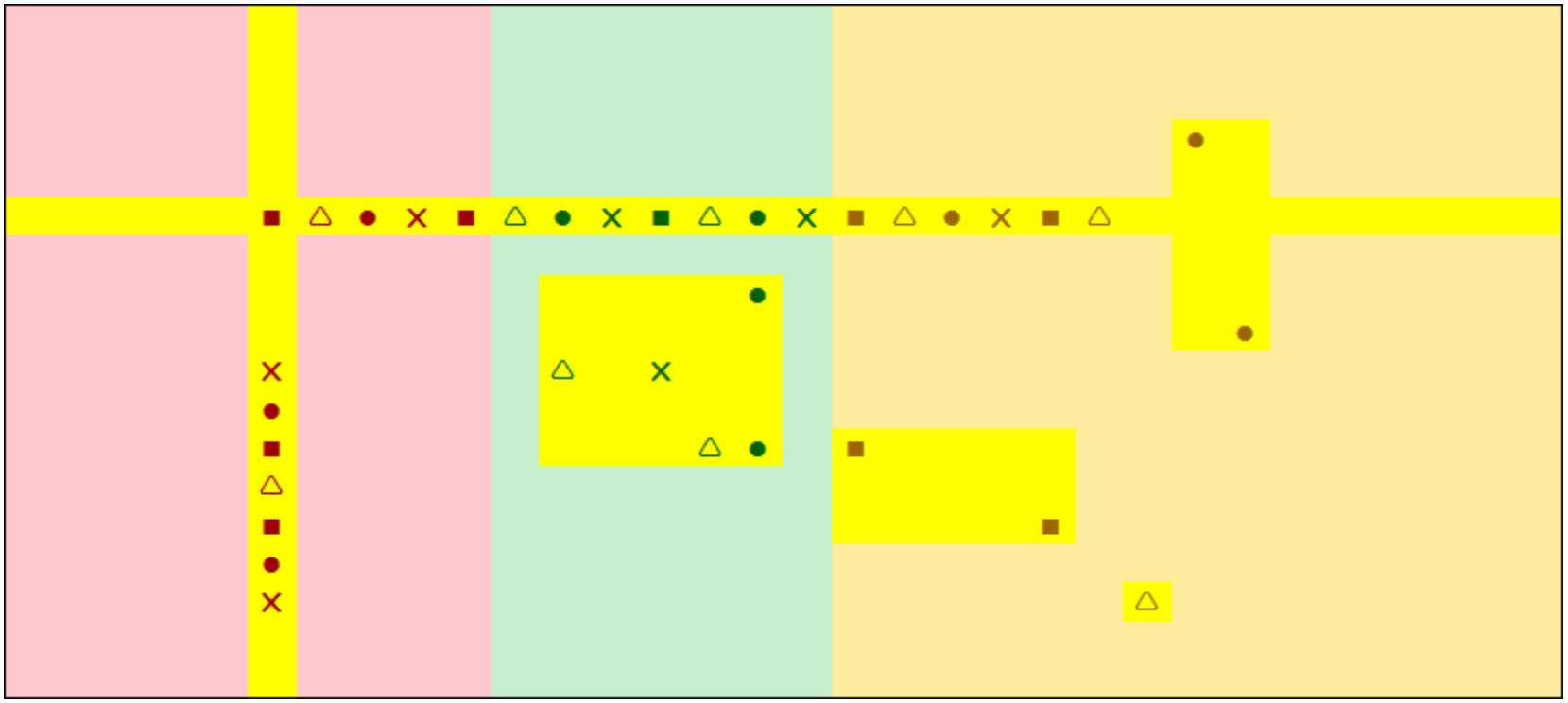
Step 4: Review formula with set C

- Are all records from set C in result?
- Review those not in the result.



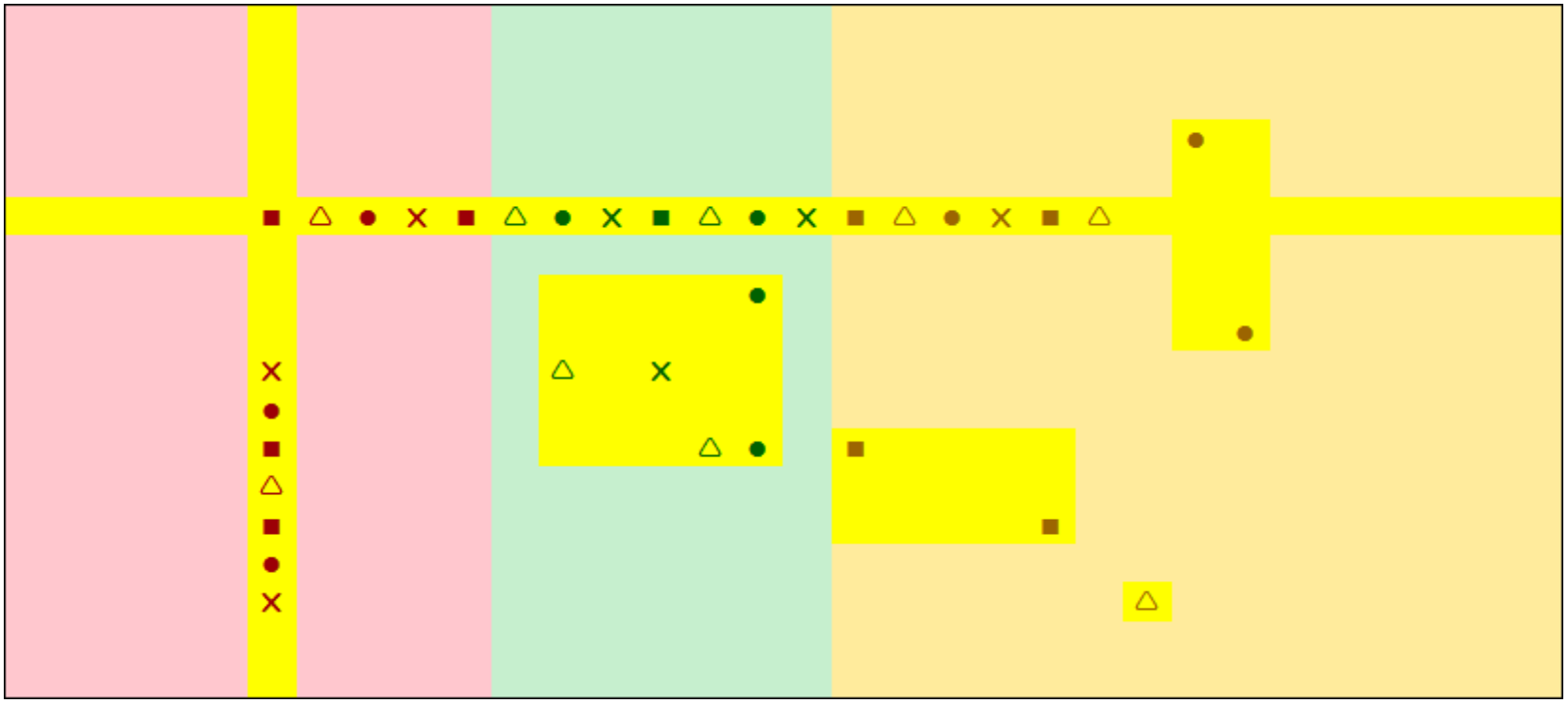
Step 4: Review formula with set C

- Revise search formula



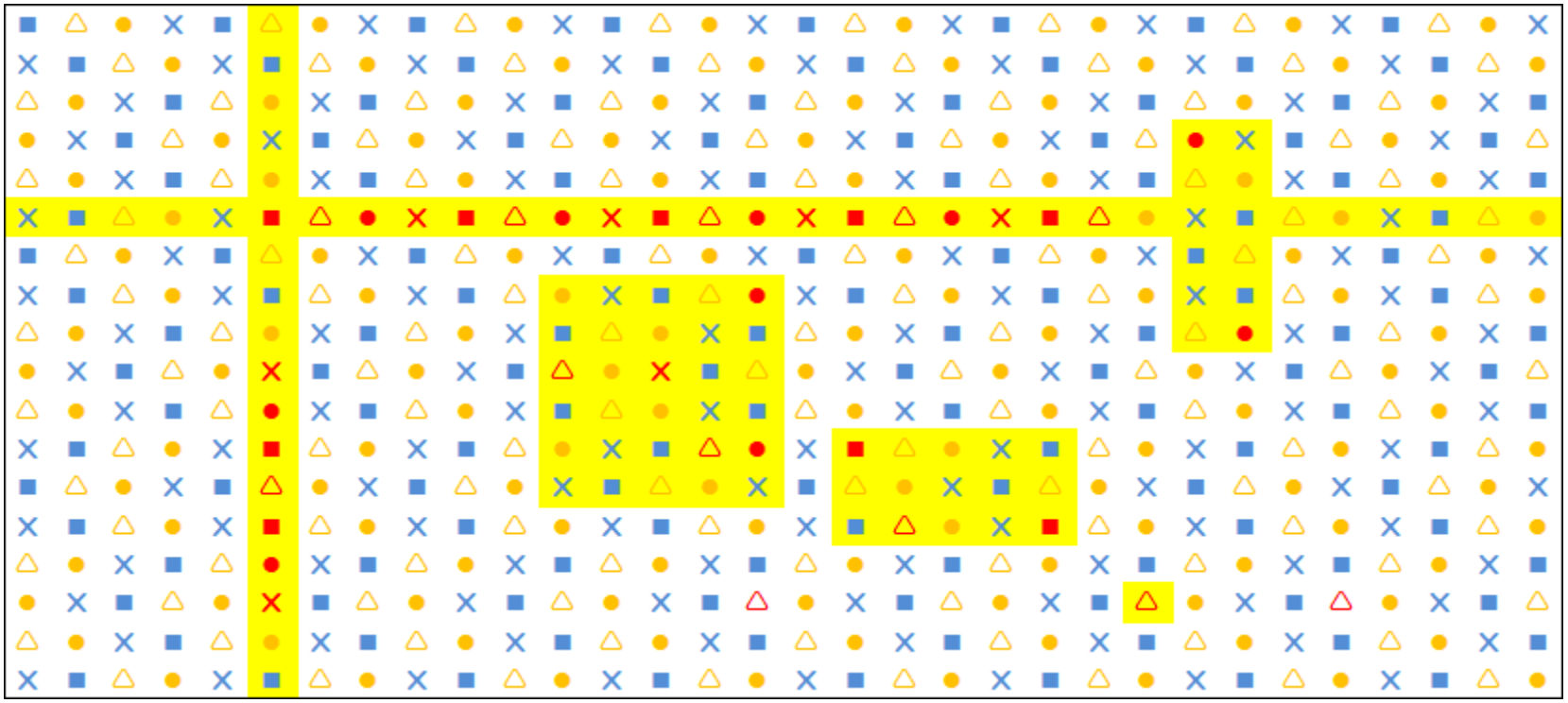
Step 4: Review formula with set C

- Apply search formula to database
- Are all records from set A, B and C in result?



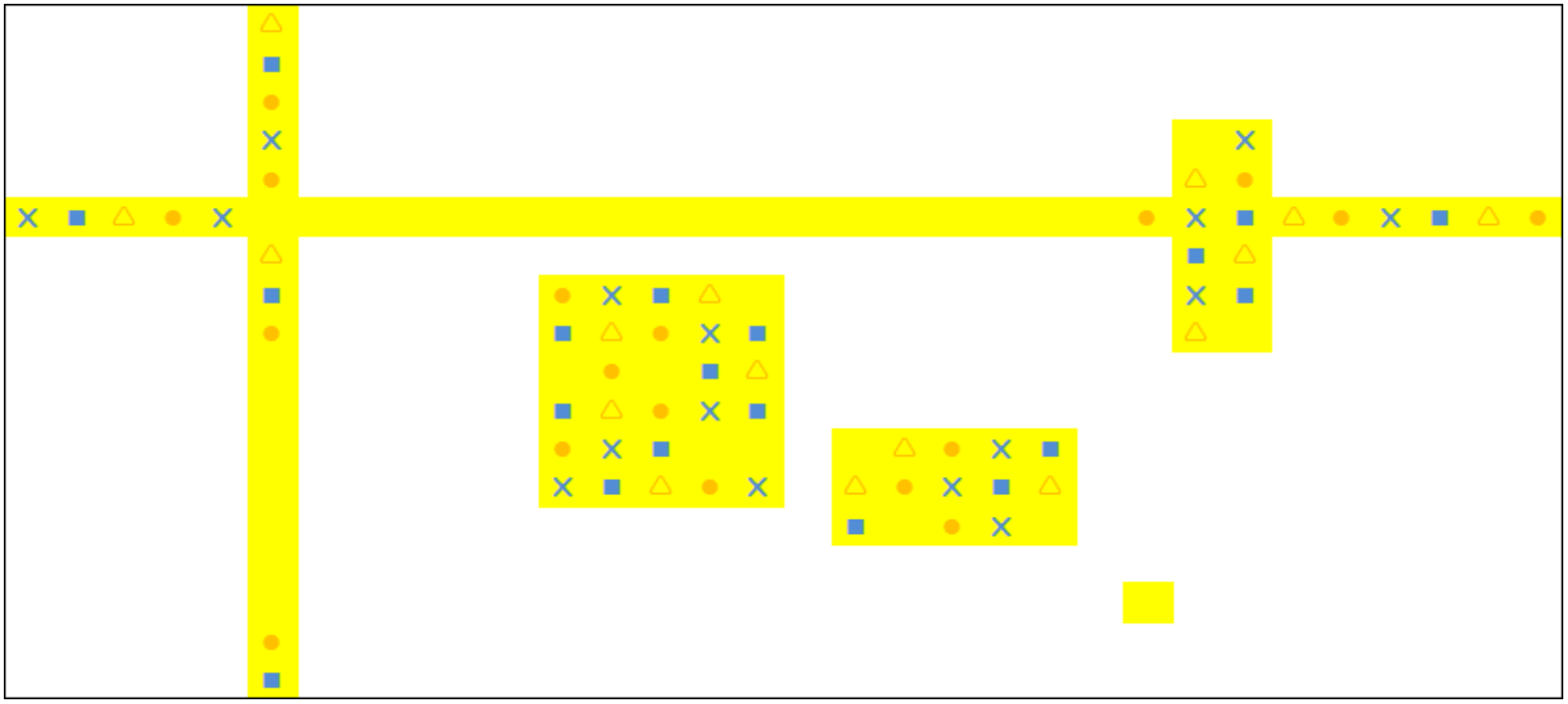
We now have a formula!

- First: Are there new ADR's detected?
- Second: Can the **Precision** be increased?



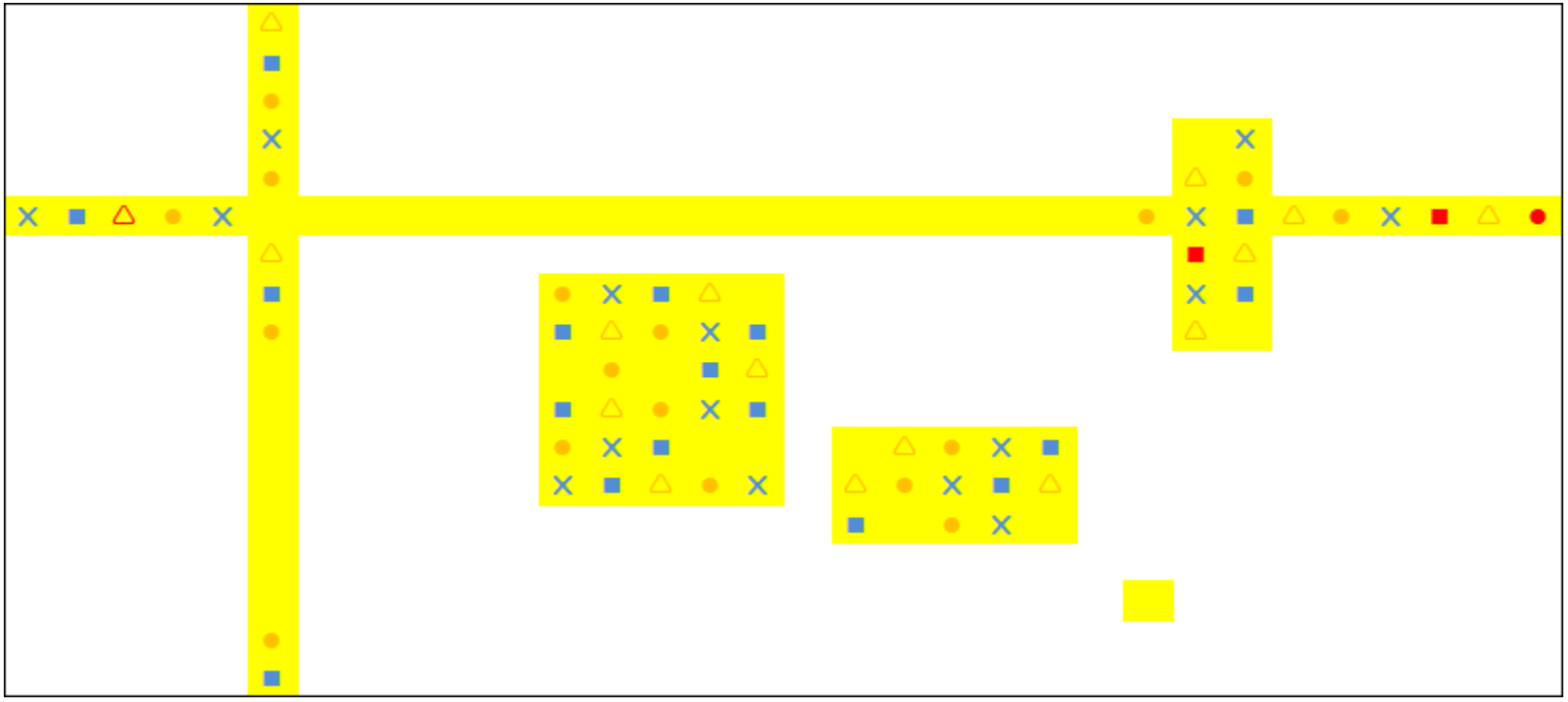
Step 5: Any new ADR's detected?

- Review all records in result not in set A, B or C.



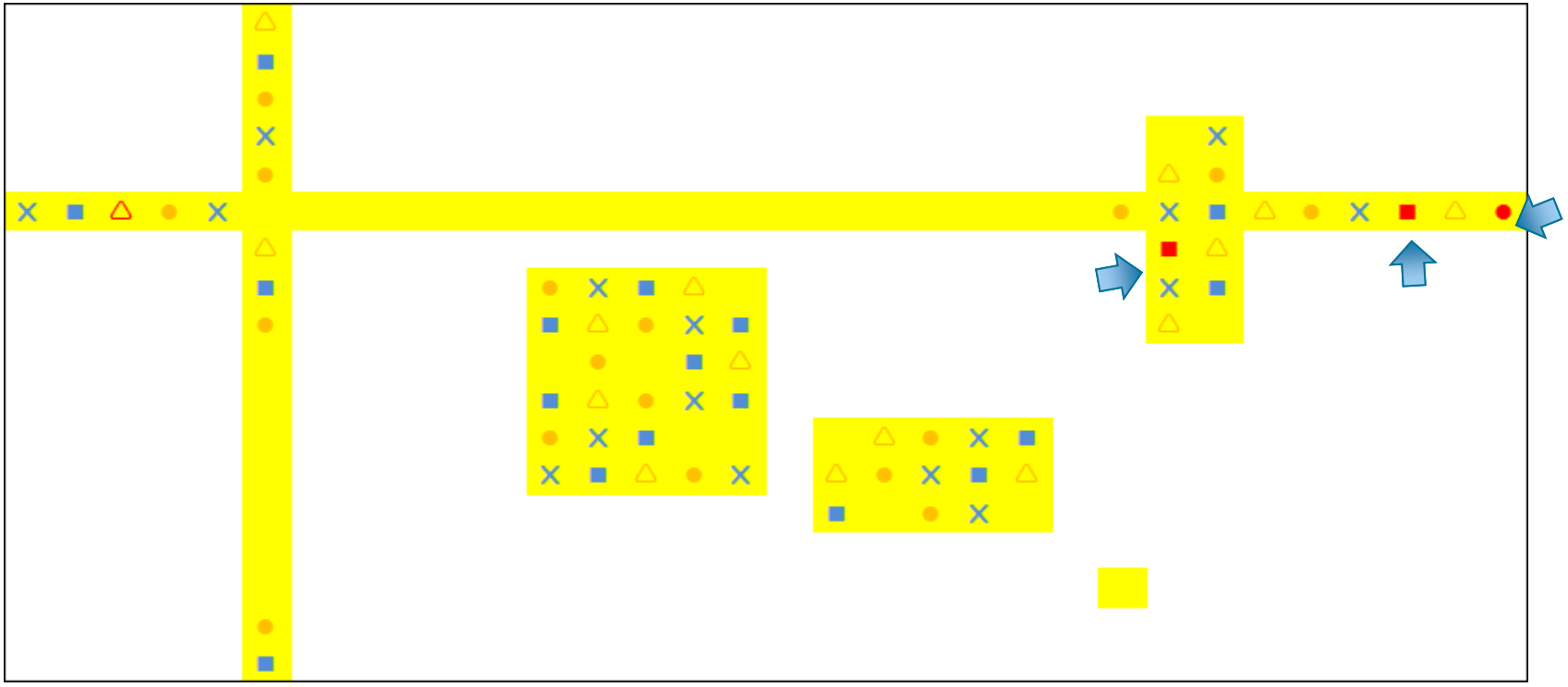
Step 5: Any new ADR's detected?

- Are there records that inform an ADR?



Step 5: Any new ADR's detected?

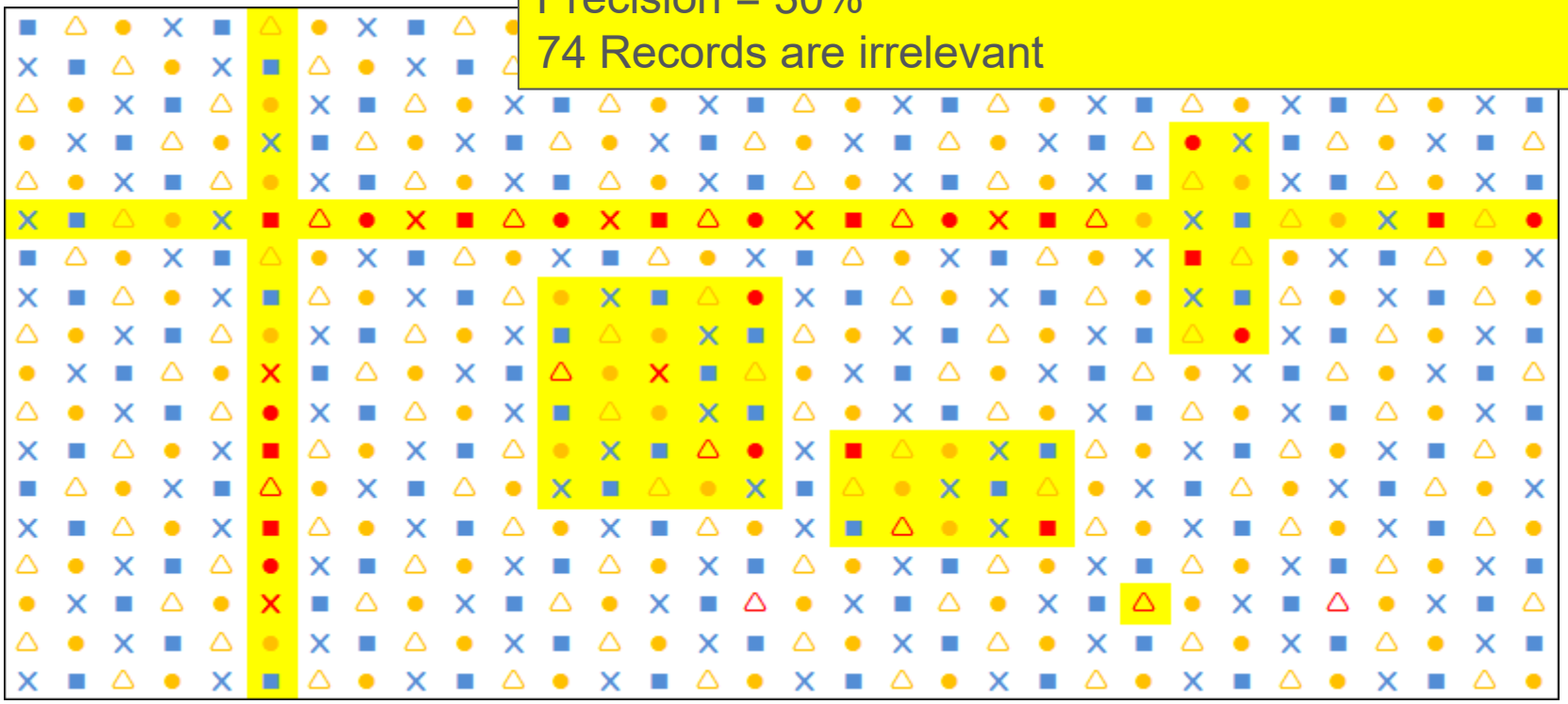
- Get the articles, review the articles, inform Safety of any positive findings



Step 6: Increase Precision

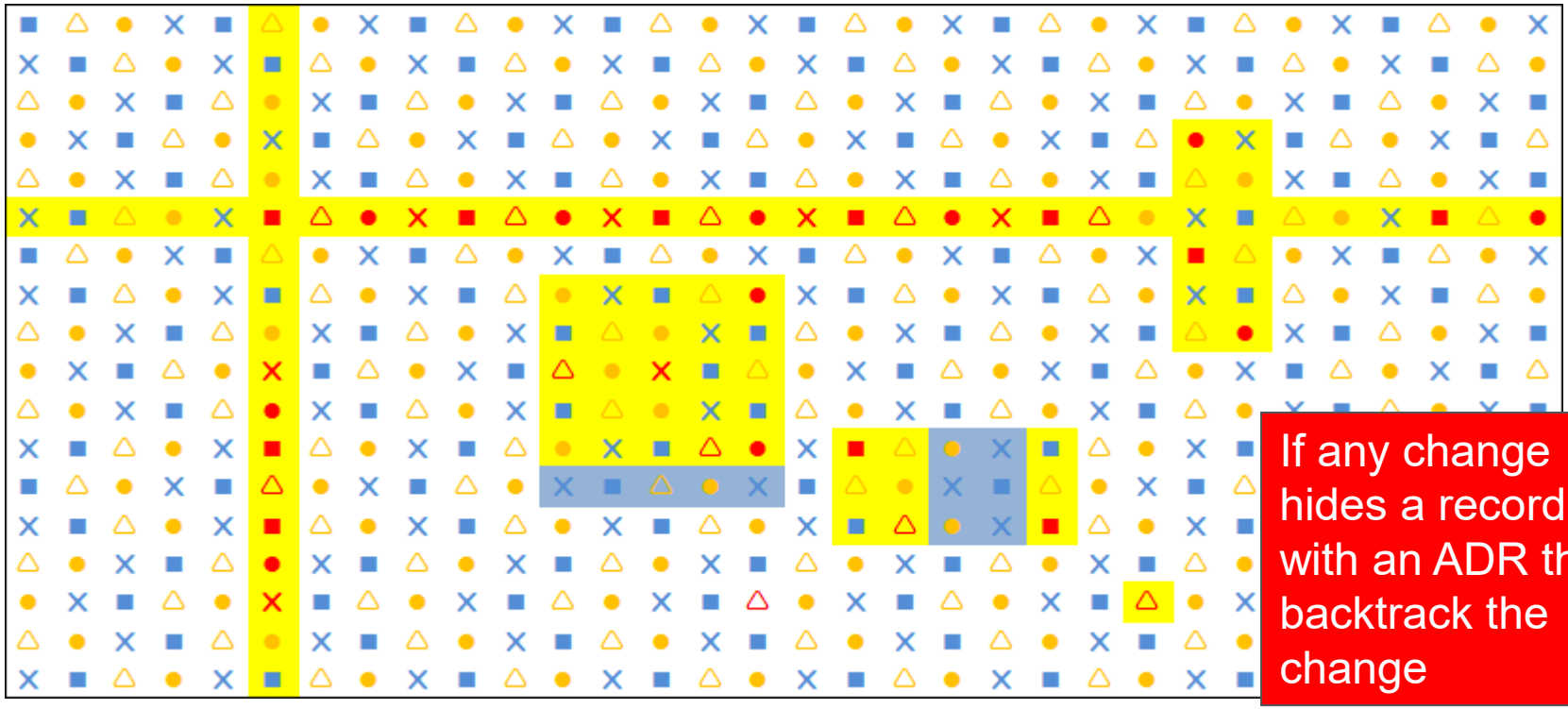
- Having achieved 100% **Recall** we may want to increase **Precision.**

This formula recovers 105 Records
31 Records contain ADR information
Precision = 30%
74 Records are irrelevant



Step 6: Increase Precision

- After reviewing some formula segments…



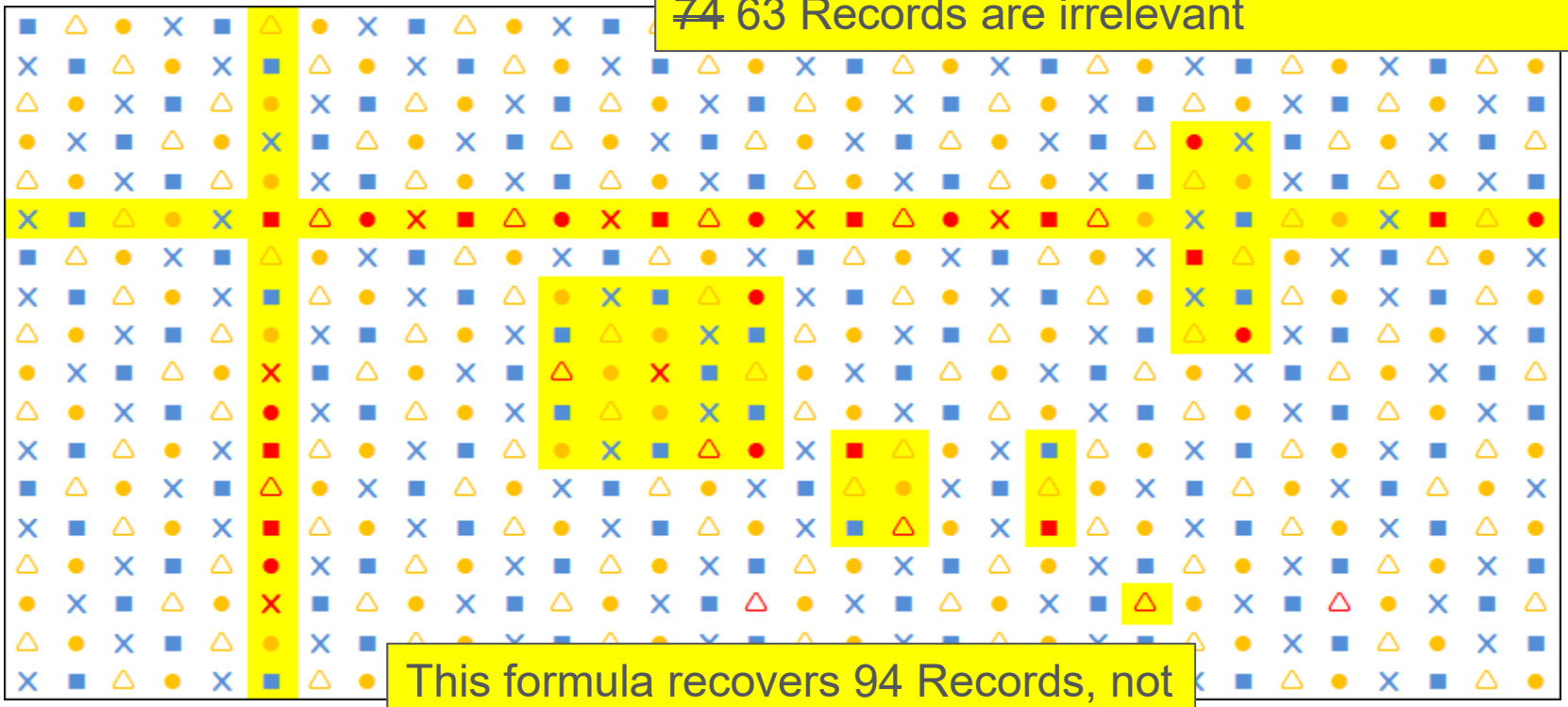
If any change hides a record with an ADR then backtrack the change

Embase®

Step 6: Increase Precision

- We can eliminate some results.

This formula recovers ~~105~~ 94 Records
31 Records contain ADR information
Precision = ~~30%~~ 33%
~~74~~ 63 Records are irrelevant

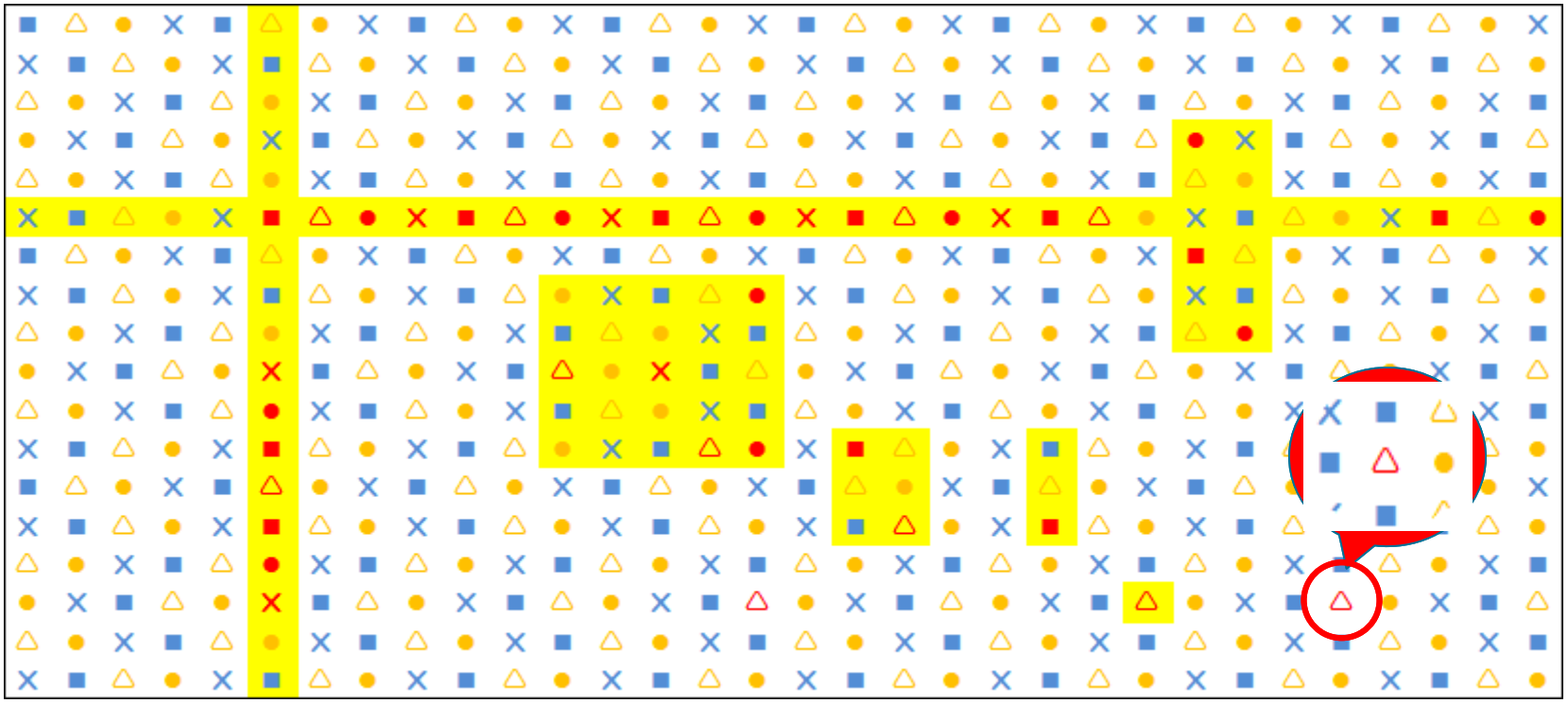


This formula recovers 94 Records, not 105 = 10% increase in productivity

Embase®

Step 7: Verify outliers

- This article has an ADR, **but the record does not**. This is result of the database not being indexed for ADRs



Step 7: Handle outliers

- Document the absence of information on the record.
- Perform **risk assessment** on the ADR:
  - Is the ADR new?
  - Is the ADR serious?
  - Was it reported by HCP?
- If the ADR/AE is NEW then, probably, this would be secondary subject of the record indexing.
  The same for seriousness.
- Document findings.

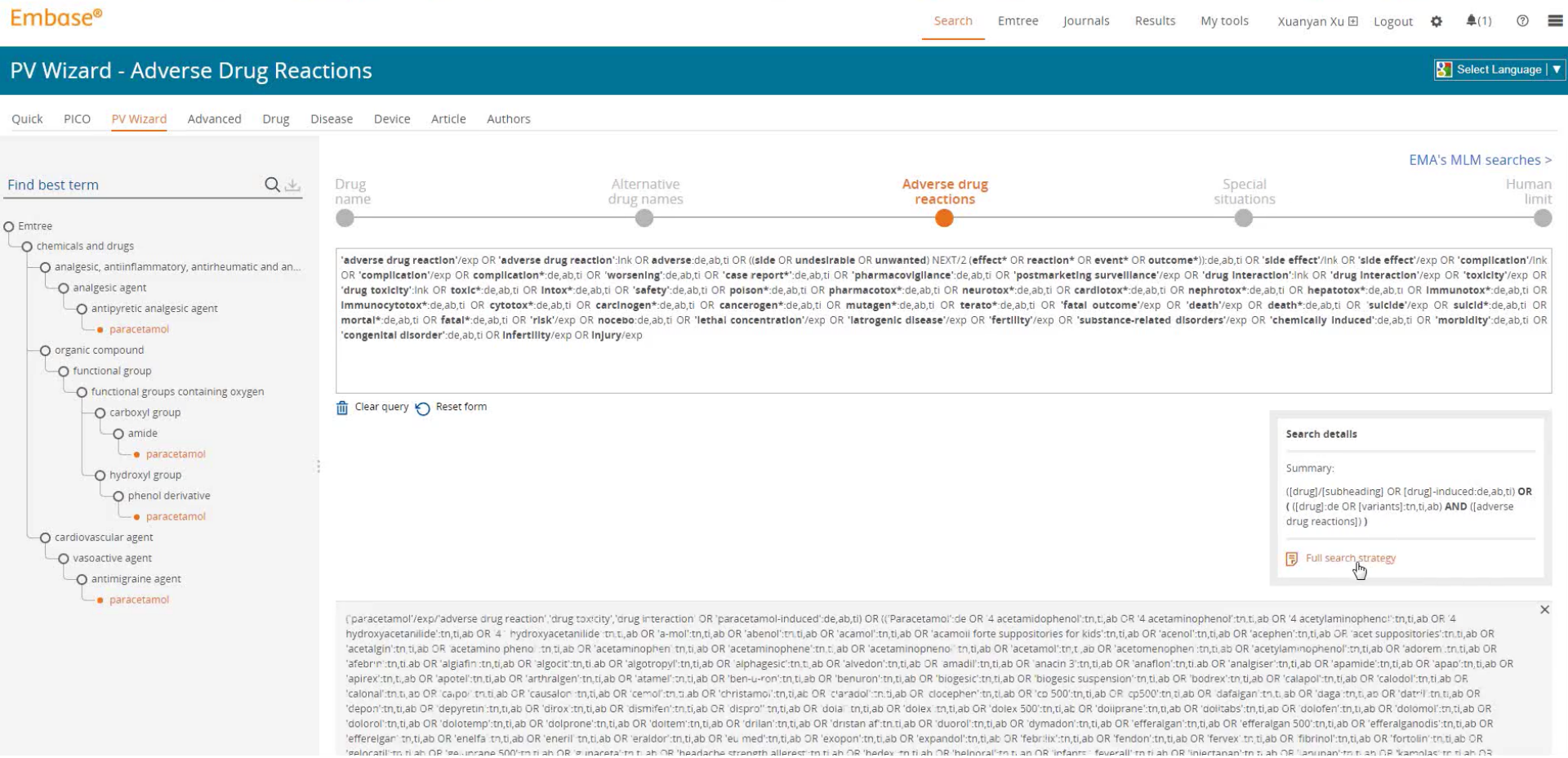
Recap: Gold set exists

- 1: Gather a Gold set
  Split into 3+ set (A,B,C)
- 2: Build a formula that retrieves all records on set A
- 3: Apply to set B, review formula to retrieve all missing records
- 4: Apply to set C, review formula to retrieve all missing records
- 5: Review other records retrieved for new ADR's
- 6: Increase **precision**, never loosing **recall**
- 7: Verify, and document, outliers

Scenario II: No Gold set exists

Scenario II: No Gold set exists

Build your own, with help from PV Wizard(*)!



Embase®

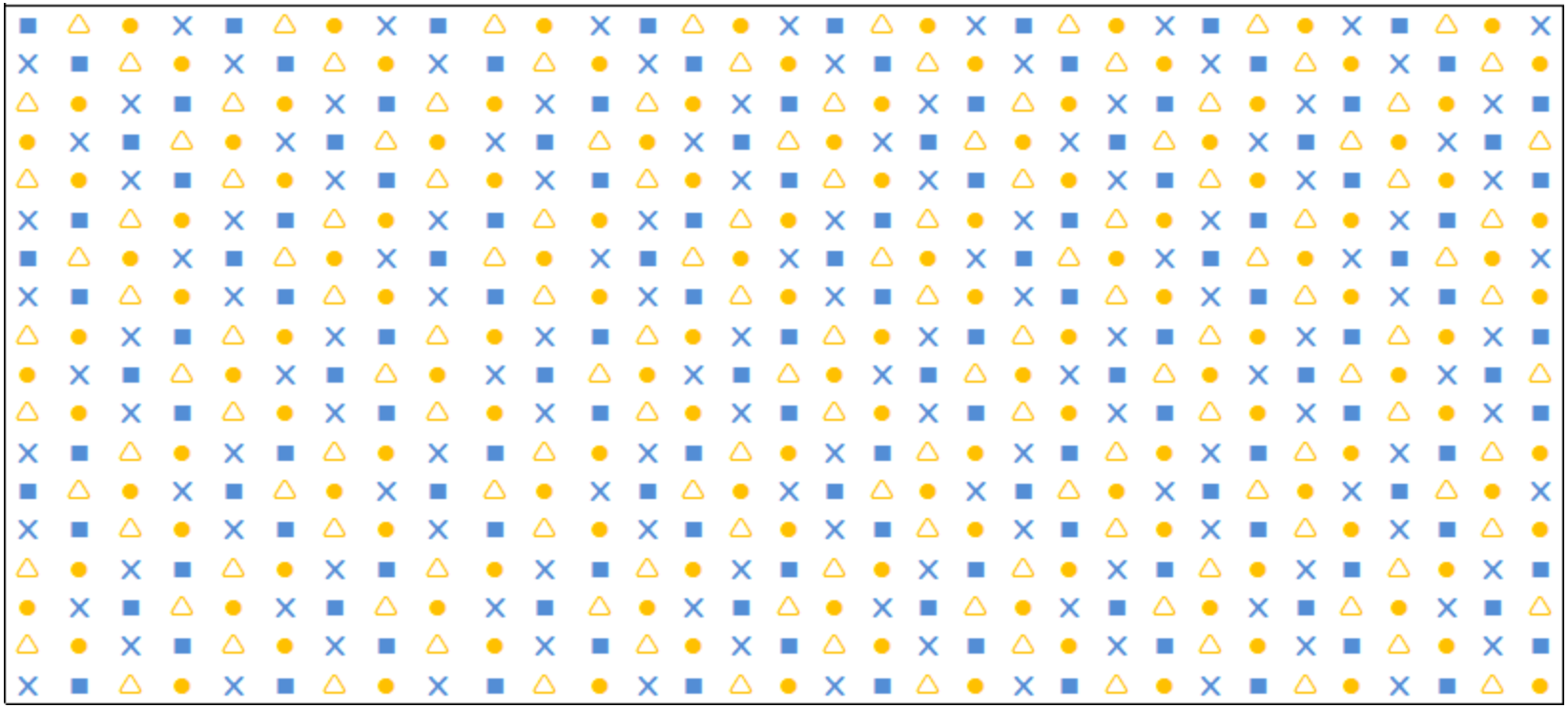
Scenario II: No Gold set exists

Build your own, with help from PV Wizard(*)!

- Execute the PV Wizard for the product
- Execute a search for all records that mention the product.
- From this EXCLUDE the PV Wizard results
- Verify everything that remains (or a sample of it)
- Change PV Wizard if necessary

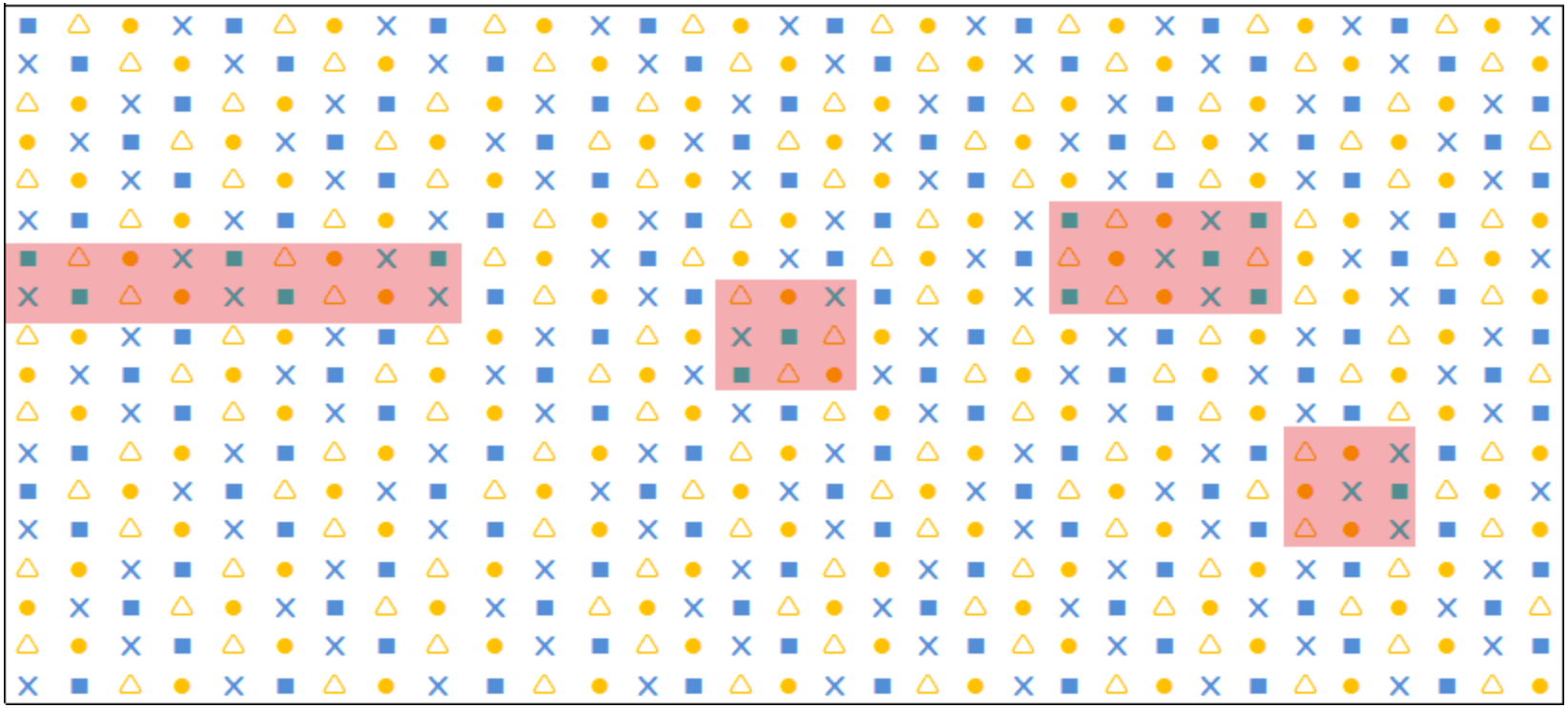
Scenario II: No Gold set exists

- For a timebox of 2 years.



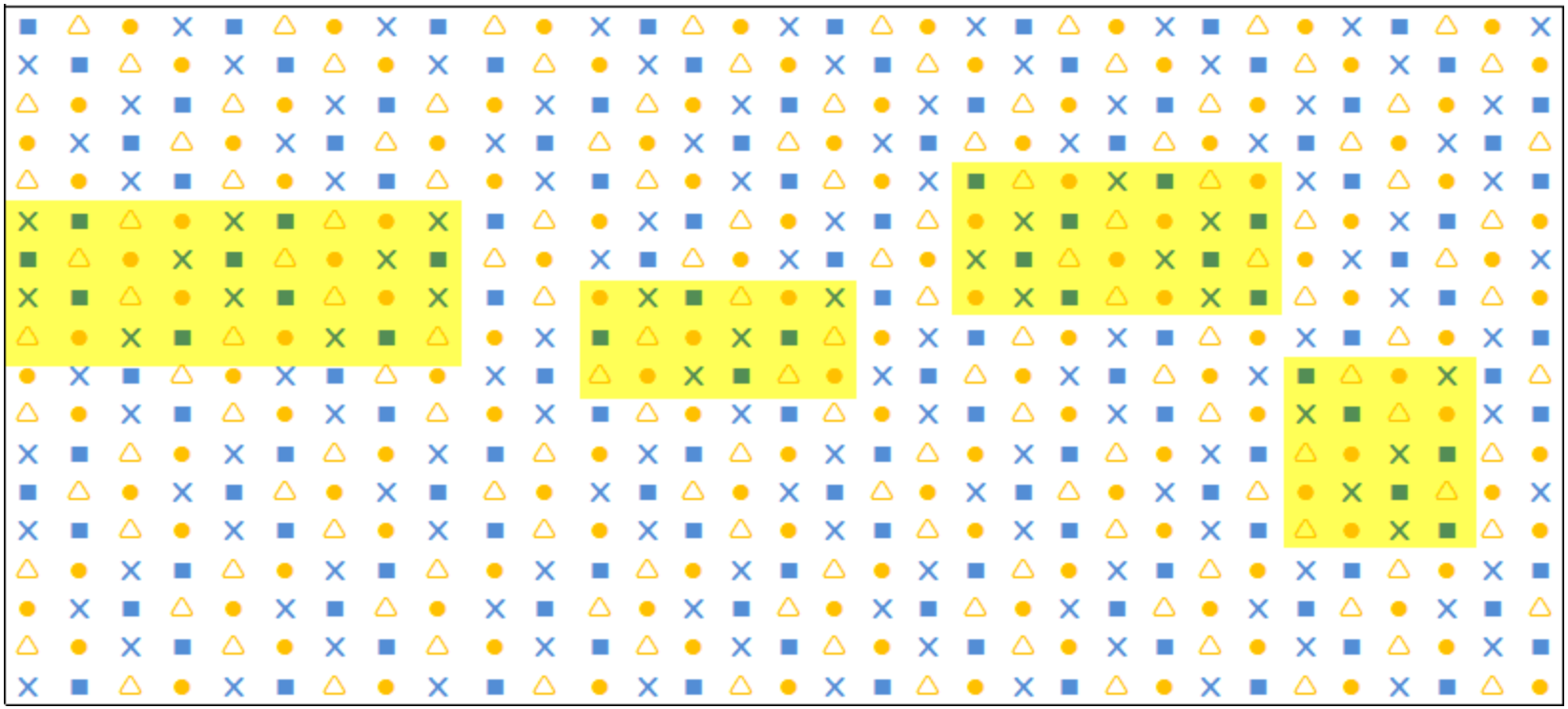
Scenario II: No Gold set exists

- For a timebox of 2 years.
- Find PV Wizard results for your product



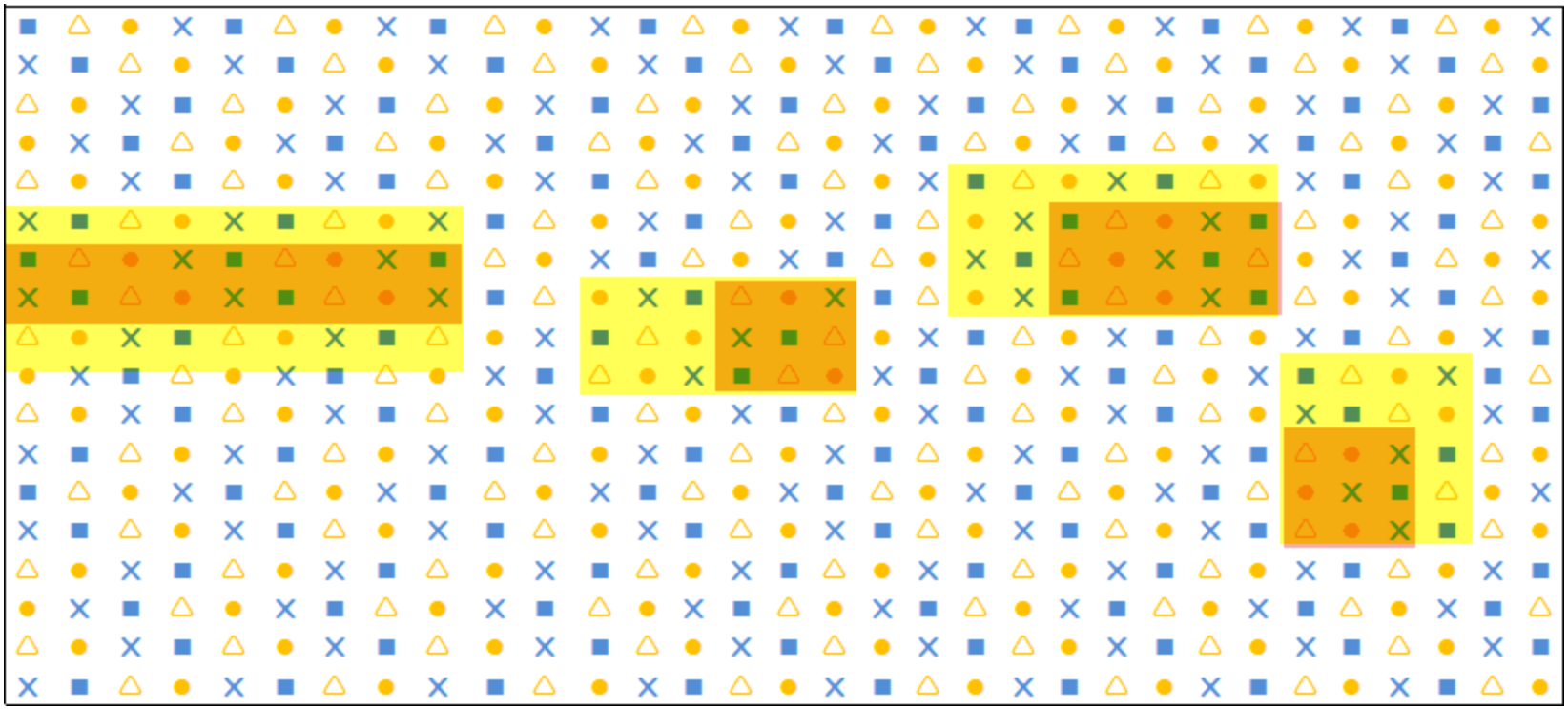
Scenario II: No Gold set exists

- For a timebox of 2 years.
- Find PV Wizard results for your product
- Find all results for your product



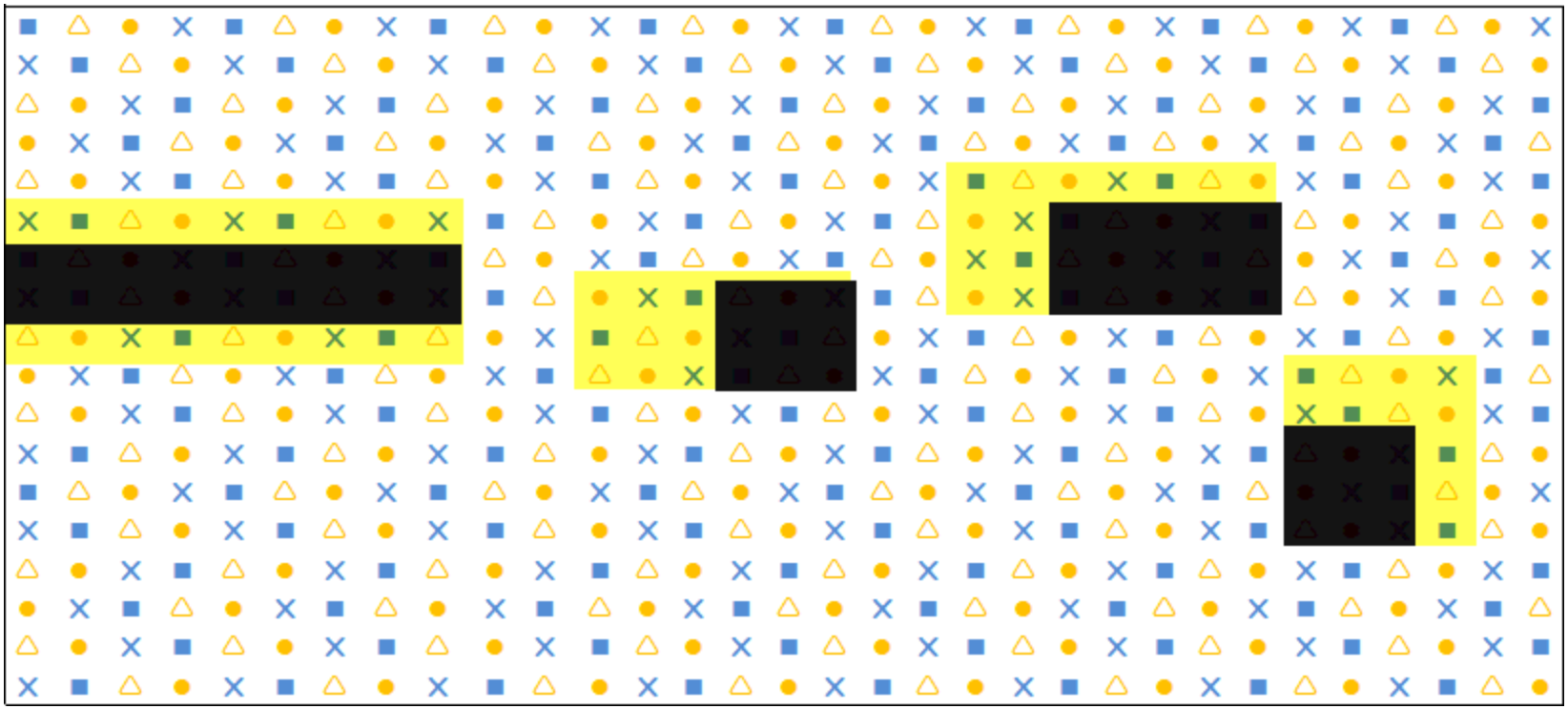
Scenario II: No Gold set exists

- Find PV Wizard results for your product
- Find all results for your product
- Find the *delta*



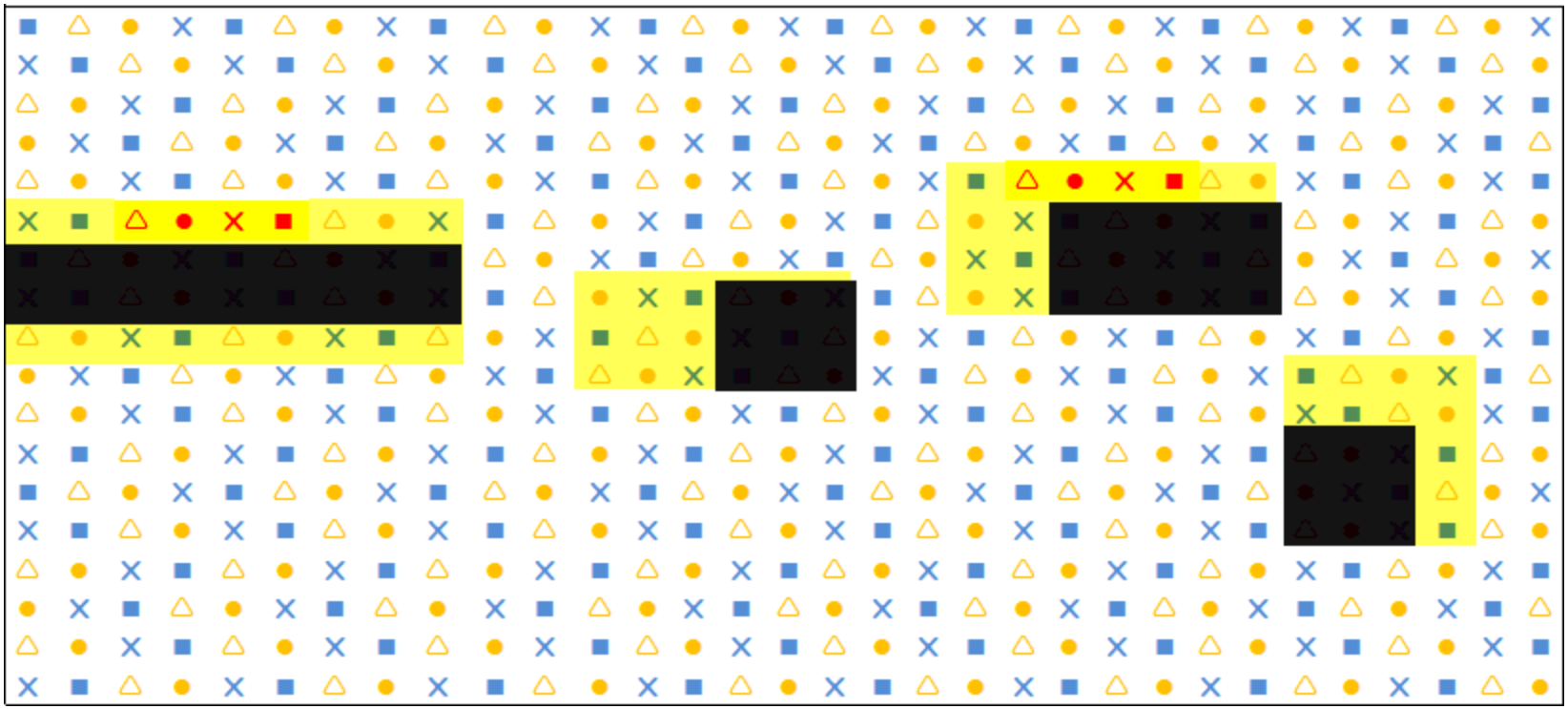
Find PV Wizard results for your product

- Review all records in the *delta*
  If the delta is too large (> 150-200) records review only a sample, ex: SQRT(*delta)*+1



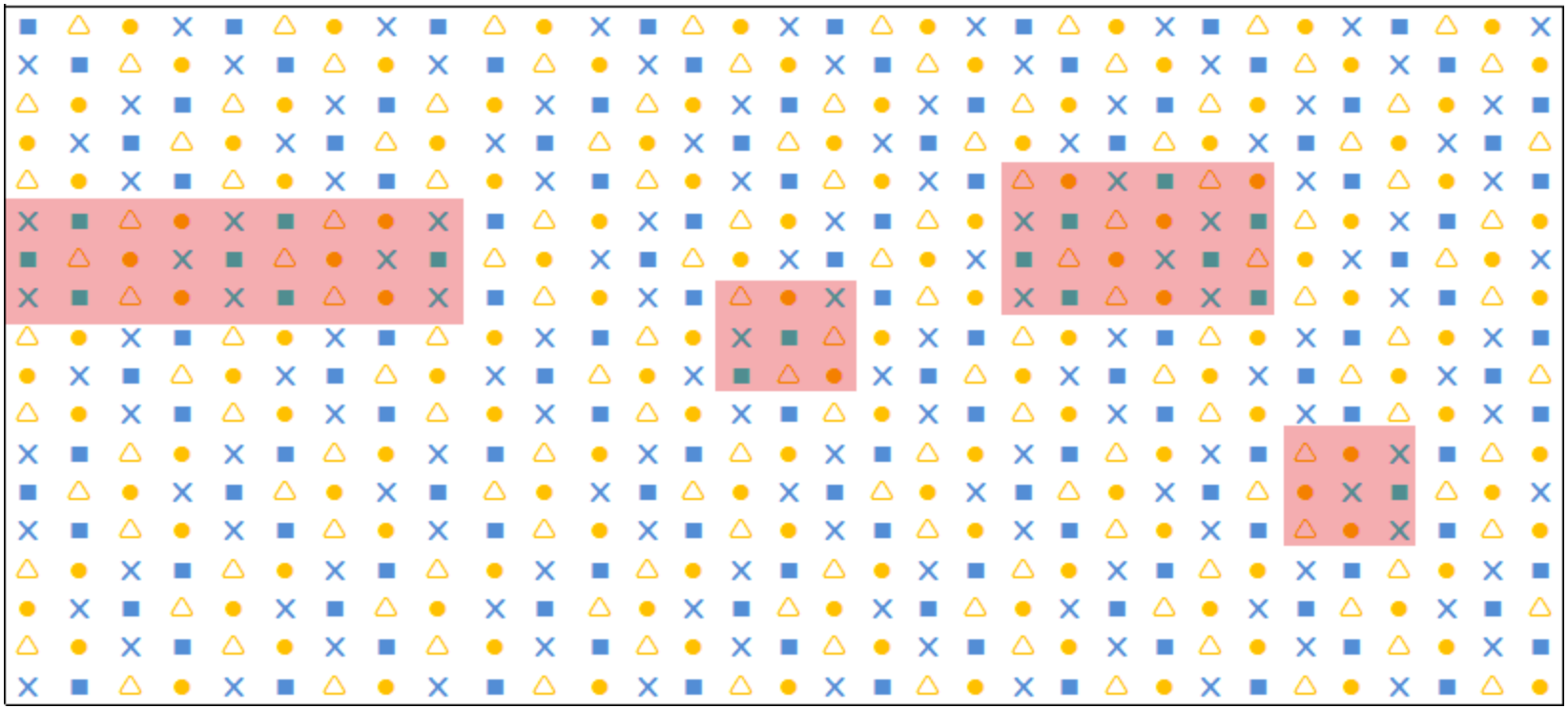
Scenario II: No Gold set exists

- If records with ADR's are found then
  Change search formula suggested by PV Wizard in order to also recover these records



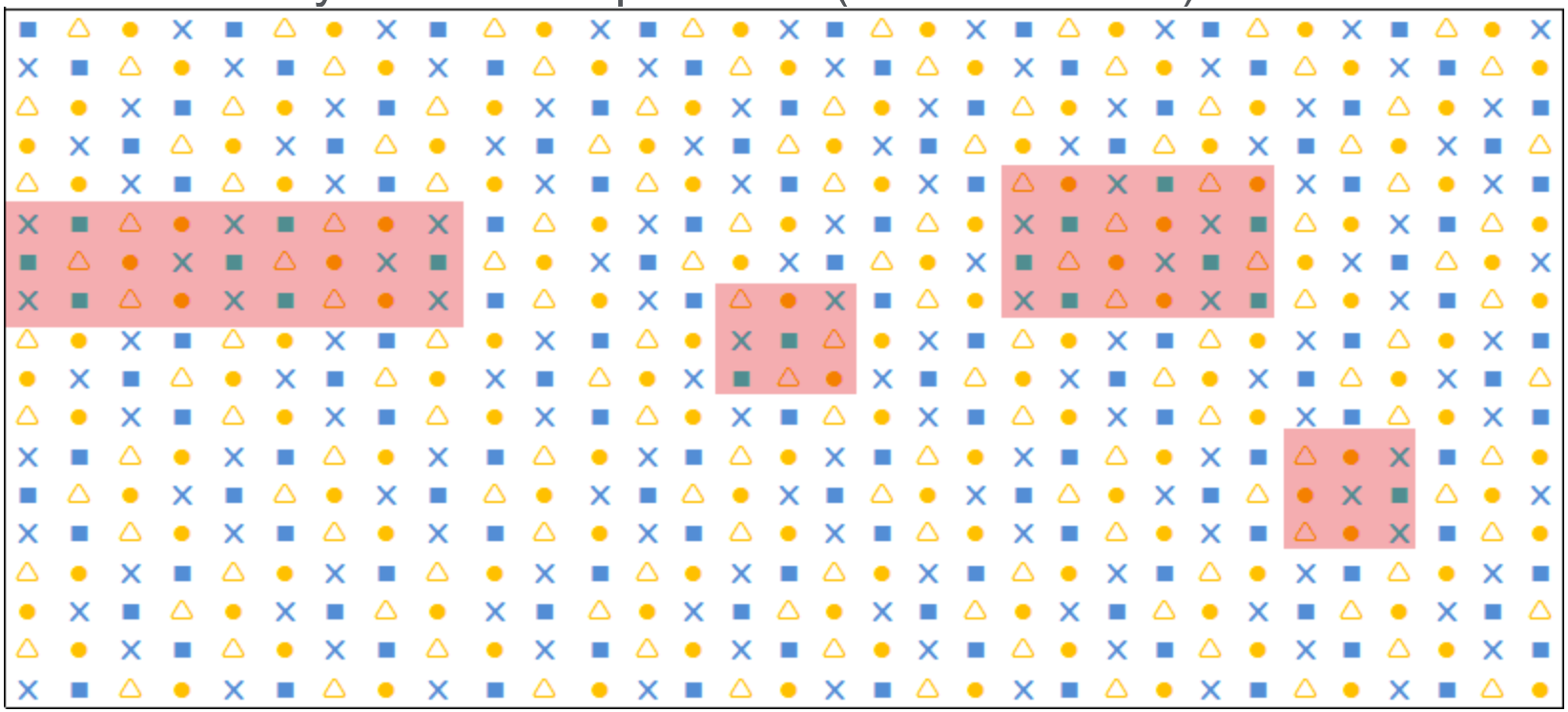
Scenario II: No Gold set exists

- If records with ADR's are found then
  Change search formula suggested by PV Wizard in order to also recover these records



Scenario II: No Gold set exists

- If records with ADR's are found then
  Change search formula suggested by PV Wizard in order to also recover these records
- You can try to increase precision (see *scenario 1*).



Scenario III: IND

SCENARIO III: Investigational New Drug

- Review all records that mention the IND using a simple formula.
- Volume of results will decide when Precision needs to be increased
- At this point a complex formula is required: proceed with the process for scenario I, as a Gold set has been collected.

Periodic Review of Search formula

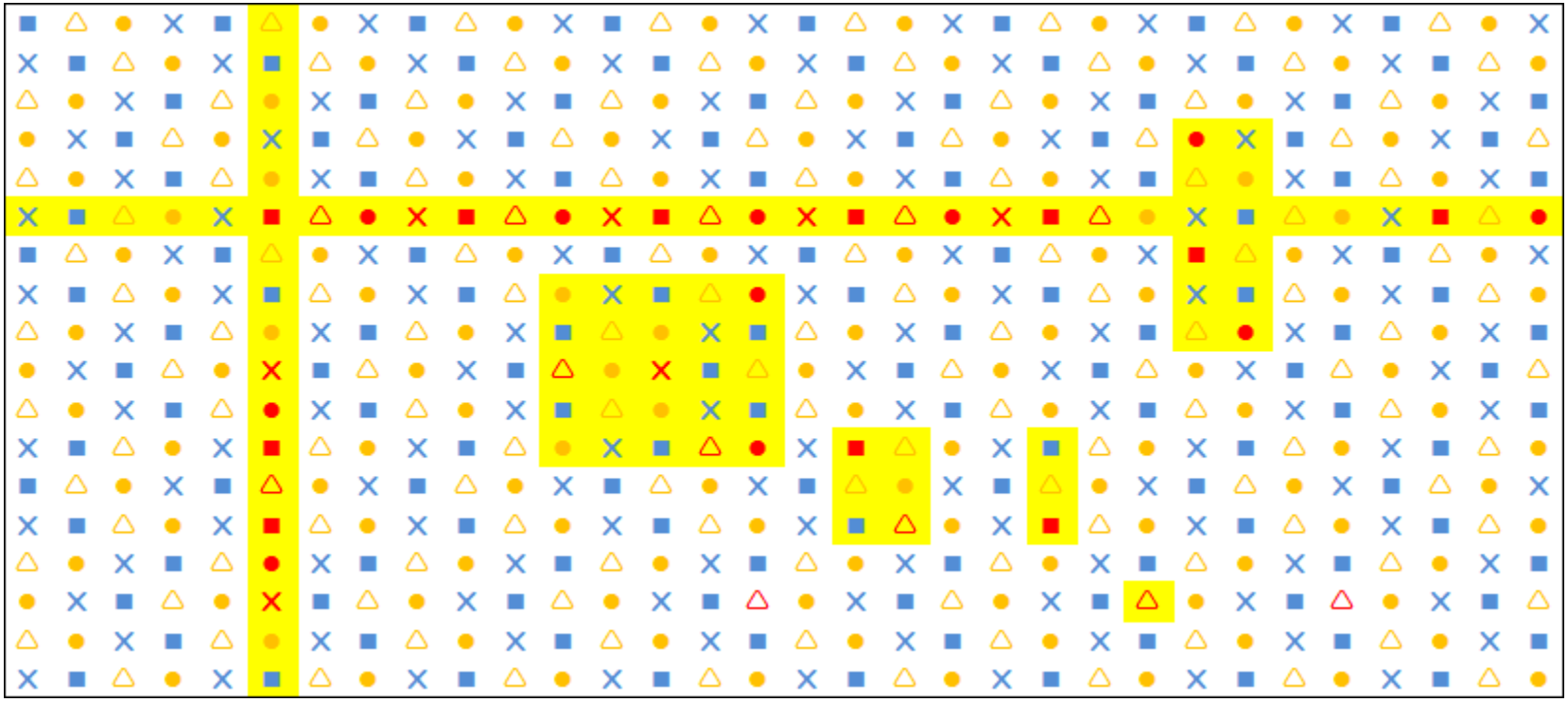
Review regularly: Why?

- A&I databases change
- New search artefacts are introduced
- New indexing policies are enforced
- *ISO 9001: Continuous Improvement*

Always perform the review maintaining *documented evidence which provides a high degree of assurance that* ***the new search formula*** *will consistently produce* ***a record set*** *meeting its pre-determined specifications and quality attributes.*

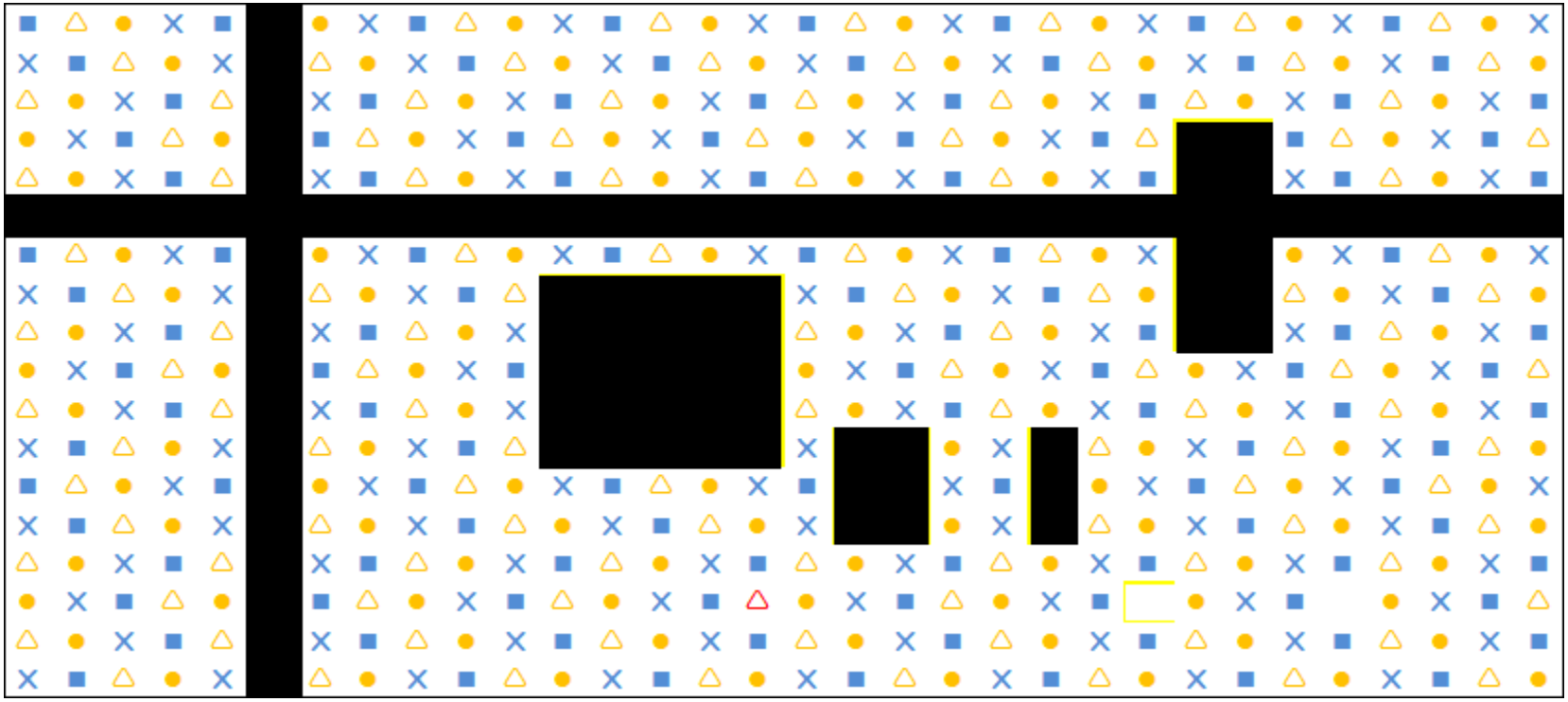
Review regularly: How?

- Option 1: Review all unexamined records



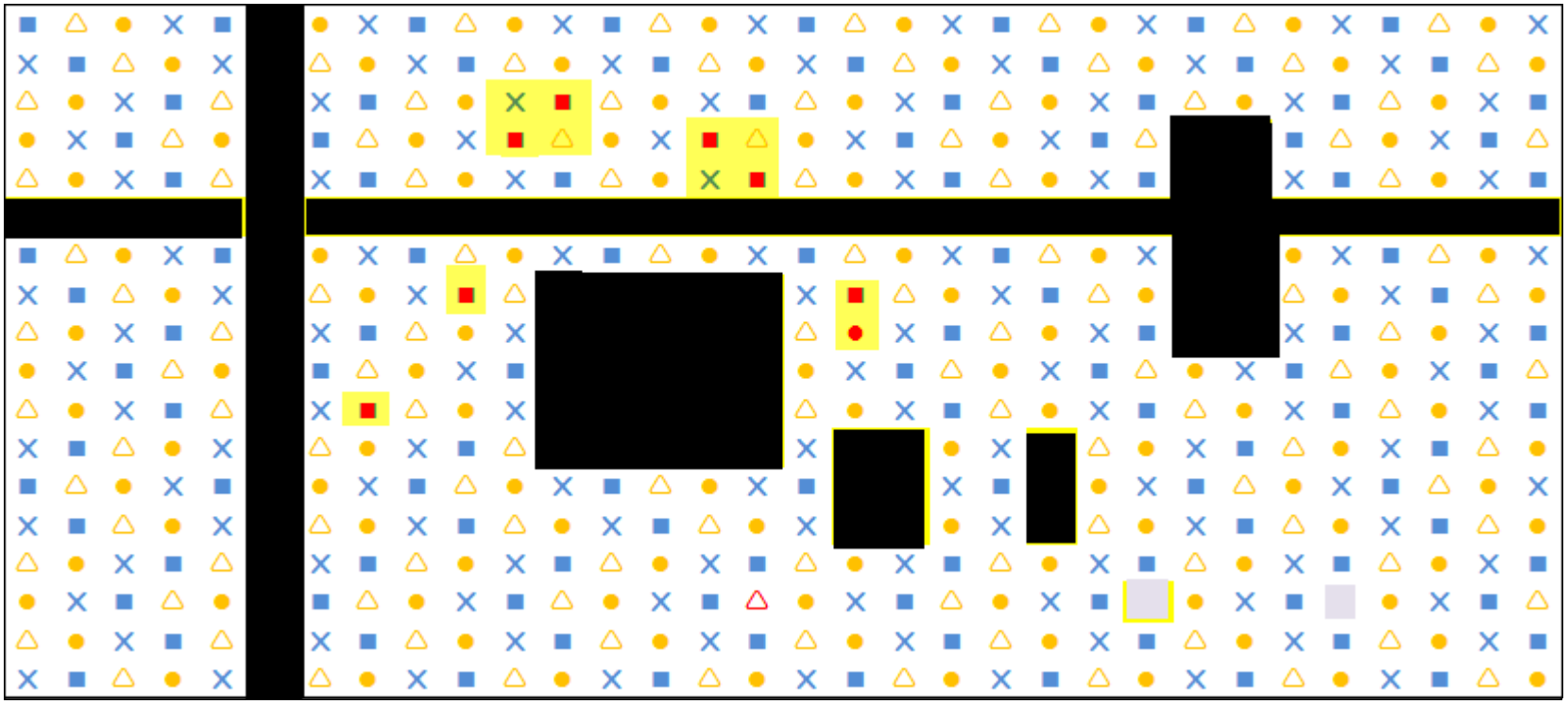
Review regularly: How?

- Option 1: Review **all** unexamined records
- Option 2: Review SQRT(*unexamined*) +1 records



Review regularly: How?

- A few outliers may be identified.
- A few new search directives may be identified.
- Add these new directives to your formula going forward.



Shortcomings

A&I databases: Shortcomings

> *because any document can, in principle, provide answers to an infinity of questions, subject analysis should establish priorities based on the specific user groups served (or specific services provided in the information ecology).*[…] *The best subject analysis is the one that makes the best prognosis of the future use of the document.*
>
> Hjørland, B. (2001).

- The indexers are not indexing for ADR detection purposes
  You can't query an A&I database for "Show me all relevant articles for PV purposes for drug X".

- If an ADR is not relevant to the main topic of the article… there may be no reference to the ADR in the record

Q & A time

For PV Wizard

See webinar:

- Best practices for pharmacovigilance & literature monitoring
  https://attendee.gotowebinar.com/recording/4680593895774109185
  July 2017

ELSEVIER

Thank you for your attention

Júlio dos Anjos
Senior Product
Development Manager
J.dosAnjos@elsevier.com
janjos

Embase®