

The Development of a New Tool for Data Quality Improvement and Use of an Immunization Registry in the 21st Century



Deborah Walker, PhD
*Citywide Immunization Registry
New York City Department of Health
e-mail: dwalker@dohlan.cn.ci.nyc.ny.us*

Vikki Papadouka, PhD
*Citywide Immunization Registry
New York City Department of Health*

Noam Arzt, PhD
*HLN Consulting, LLC
e-mail: arzt@hln.com*

Presented at:

***National
Immunization
Conference***

*34th Meeting
July 5 - 8, 2000
Washington, DC*

Background

New York Citywide Immunization Registry (CIR)

- ◆ Mandatory reporting started January 1, 1997 by Amendment to New York City Health Code
- ◆ All providers must report immunizations given to children under age 8
- ◆ 95% of institutions and 86% of private providers are currently reporting
- ◆ 100% of institutions and 92.5% of private providers have ever reporting
- ◆ Contains records for 1.8 million children with over 9.9 million immunizations
- ◆ Population based; Vital Records loaded every 2 weeks
- ◆ De-duplicating of records a high priority

Objectives

To describe a new query tool
“Smart Search” which
searches the CIR for possible
matching records to an input
file of children.

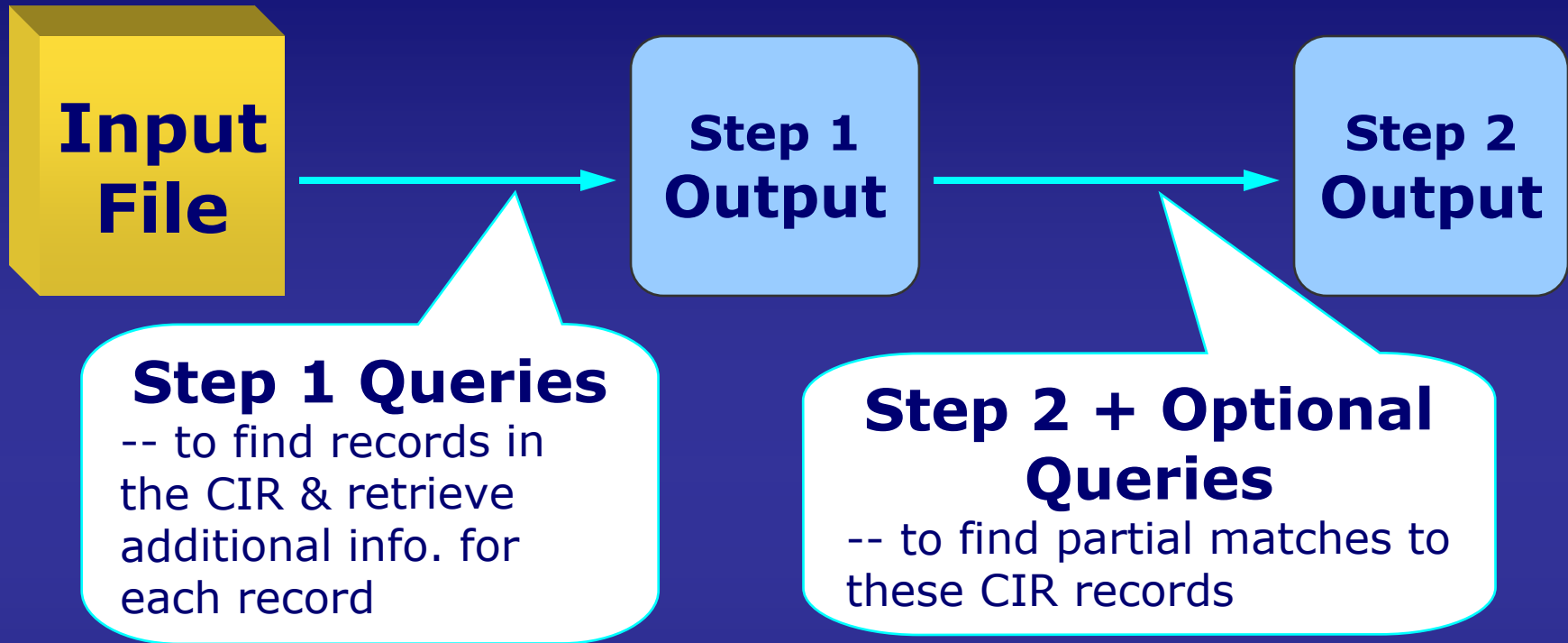
Smart Search

- ◆ A menu-driven query program to search CIR for possible matches
- ◆ User can select from over 1400 different combinations of search queries
- ◆ Used on Windows 95/98/NT desktops with log on access to CIR
- ◆ Developed in collaboration between NYC-DOH and HLN Consulting, LLC.
- ◆ Written in modules; additional functions and components easily added

Smart Search Functions

- ◆ Searches CIR for possible matches to an input file of records
- ◆ Creates file of possible matching pairs to input to **MEDD** (de-duplication program)
- ◆ Creates file to input to **Recall Generator** (reminder/recall program)
- ◆ Creates file to input to **DEI** (data exchange interface program) to exchange data with MCO's and NYC's Lead Quest Registry
- ◆ Generates patient lists and ASCII files of CIR records

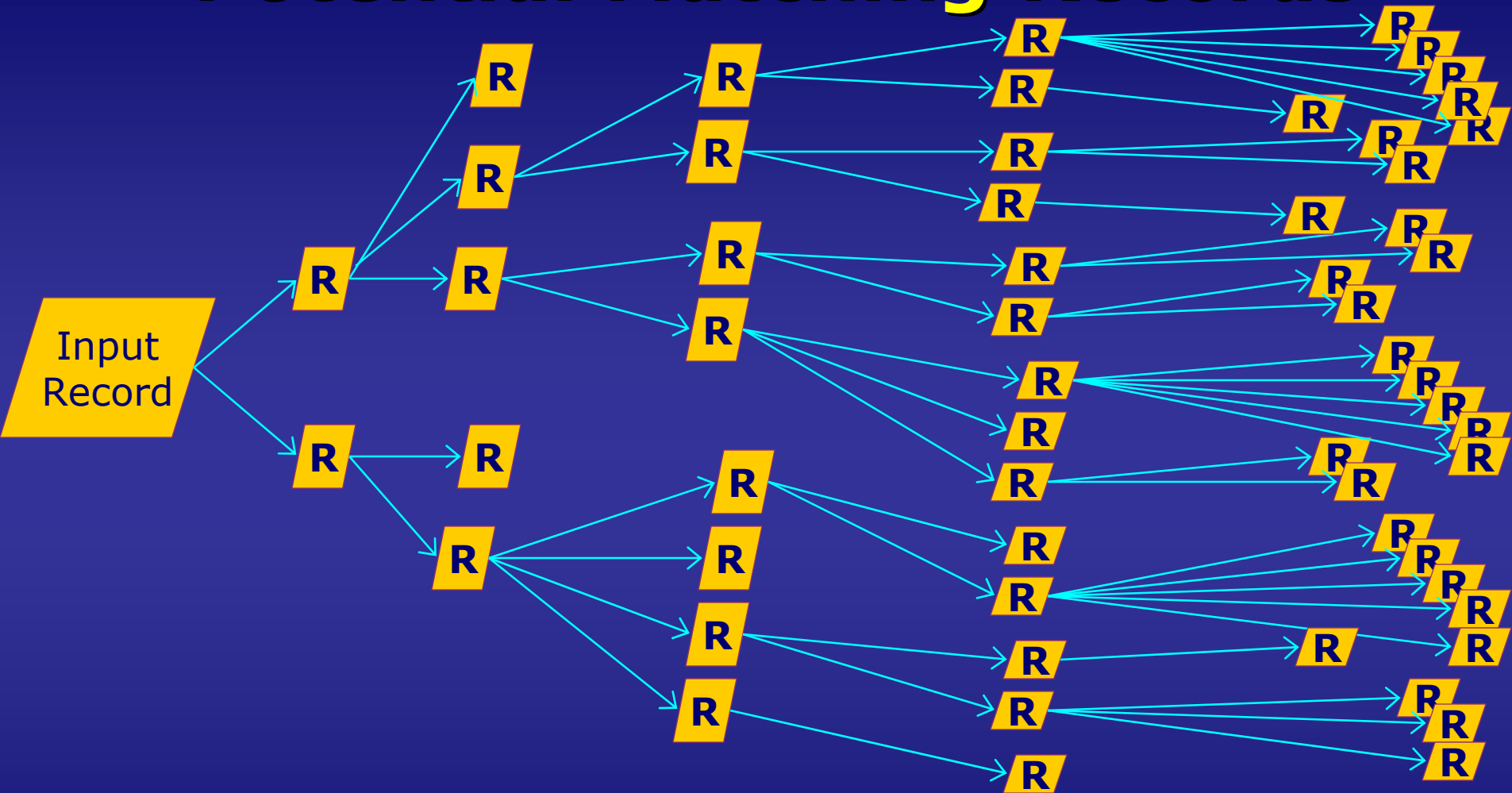
Smart Search Process



If iterative searching,

Process repeats Step 2 + Optional Queries up to 5 times.

Potential Matching Records



For every single record in Input File, multiple possible matches can be found depending upon search criteria.

Step 1 Output

Input file:

Papadouka, Vikki, 01/01/1999, F

Step 1 Output file:

```
248513512,"PAPADOUKAS","UASSILIKI",01/01/1999,F,"UP02135A",,"123",,"",,"",  
,"",,"",,"ALEXANDRA",01/01/1930,1111111111,N,N
```

Step 2 Output file:

```
248702951,"PAPAD","UASSILIKI",01/01/1999,F,"",,"",,"",,"",,"",,"",,"",,"",,"",N,N  
248702387,"PAPADOUKA","UASSILIKI",01/13/1999,F,"",,"",,"",,"",,"",,"",,"",,"",,"",N,N  
248702736,"PAPADOUKA","BABY GIRL",01/01/1999,F,"",,"",,"",,"",,"",,"",,"",,"",,"",N,N  
248702512,"PAPADOUKA","VIKKI",01/01/1998,F,"",,"",,"123",,"",,"",,"",,"",,"",,"",N,N  
248702624,"MATTHEWS","UASSILIKI",01/01/1999,F,"UP02135A",,"123",,"",,"",,"",,"",  
,"",,"",,"PAPADOUKA",01/01/1930,N,N  
248703063,"PAP","UASSILIKI",01/01/1999,F,"",,"",,"",,"",,"",,"",,"",,"",,"",N,N  
248791375,"SPENCER","SHIRLEY",02/09/1999,F,"",,"",,"123",,"",,"",,"",,"",,"",,"",N,N  
119145036,"WALKER","NATALIE",05/31/1993,F,"",,"",,"123",,"",,"",,"",,"",,"",,"",718  
6360881,N,N  
46370687,"HUIE","SHIRLEY",09/30/1996,F,"",,"",,"123",,"",,"",,"",,"",,"",,"",N,N
```


Step 1 Output

Input file:

Papadouka, Vikki, 01/01/1999, F

CIR Pairs file:

248513512 - 248702951
 248513512 - 248702387
 248513512 - 248702736
 248513512 - 248702512
 248513512 - 248702624
 248513512 - 248703063
 248702951 - 248791375
 etc.

Step 1 Output file:

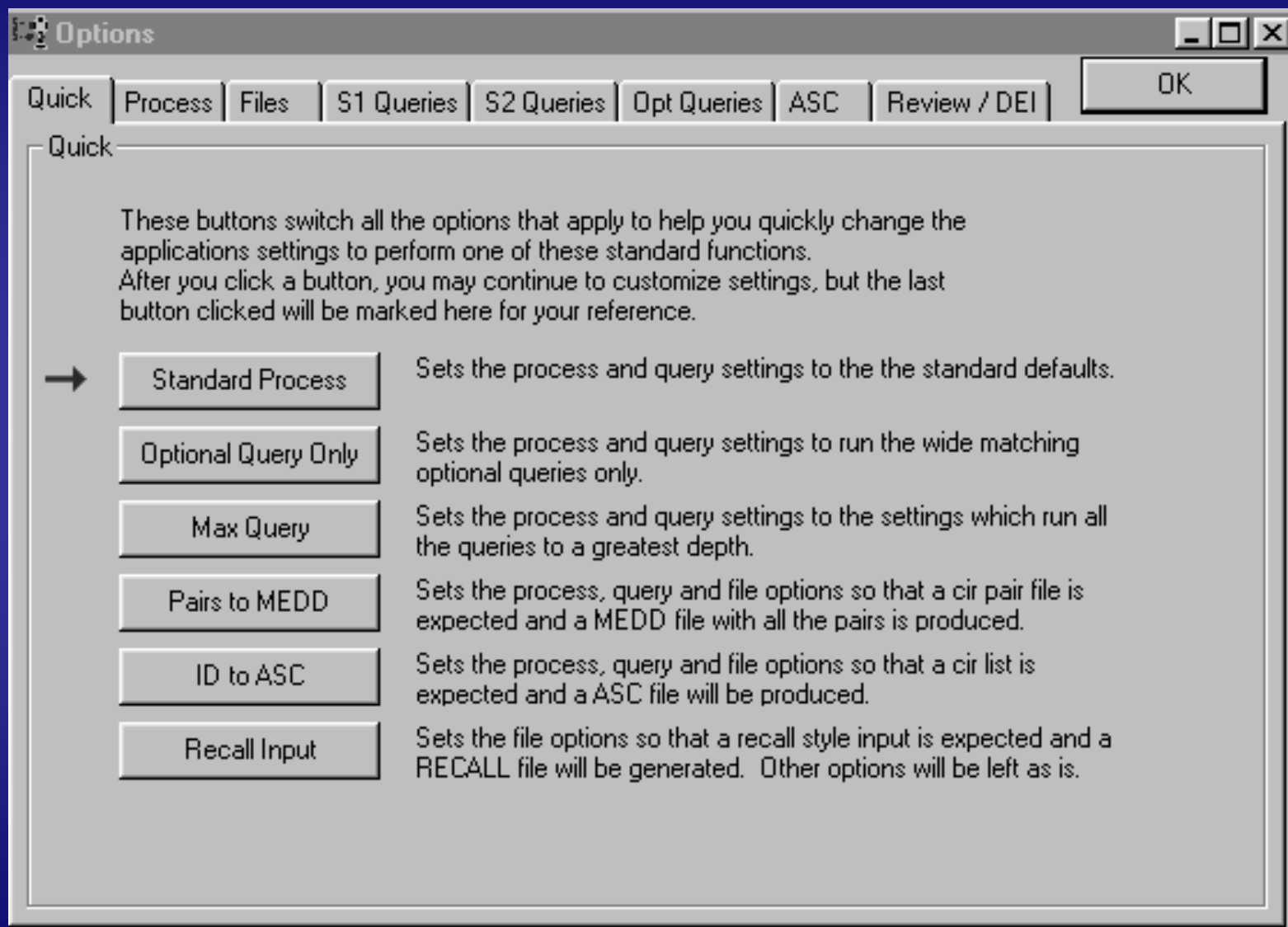
248513512
 PAPADOUKAS
 VASSILIKI
 01/01/1999
 F
 VP02135A
 123
 ALEXANDRA
 01/01/1930

248513512

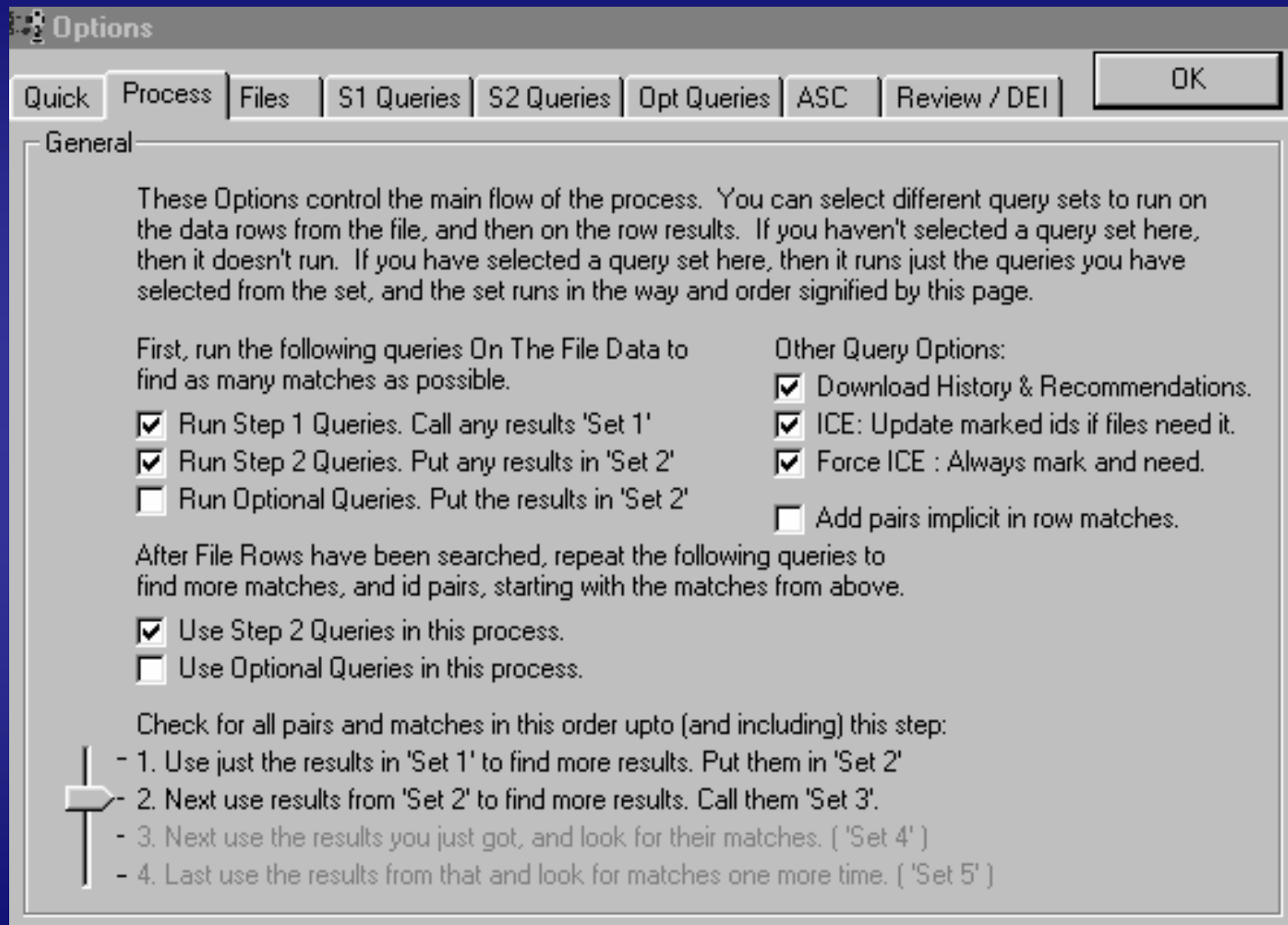
Step 2 Output file:

248702951 PAPAD VASSILIKI 01/01/1999 F	248702387 PAPADOUKA VASSILIKI 01/13/1999 F	248702736 PAPADOUKA BABY GIRL 01/01/1999 F	248702512 PAPADOUKA VIKKI 01/01/1998 F 123	248702624 MATTHEWS VASSILIKI 01/01/1999 F VP02135A 123 PAPADOUKA 01/01/1930	248703063 PAP VASSILIKI 01/01/1999 F	248791375 SPENCER SHIRLEY 02/09/1999 F 123	119145036 WALKER NATALIE 05/31/1993 F 123 7186360881	46370687 HUIE SHIRLEY 09/30/1996 F 123
--	--	--	---	---	--	---	--	---

Smart Search Functions



Process

The image shows a software dialog box titled "Options" with a standard Windows-style title bar. It features a tabbed interface with tabs for "Quick", "Process", "Files", "S1 Queries", "S2 Queries", "Opt Queries", "ASC", and "Review / DEI". The "Process" tab is currently selected. An "OK" button is located in the top right corner. The main content area is titled "General" and contains several paragraphs of text and a list of checkboxes. The text explains that these options control the main flow of the process, allowing users to select different query sets to run on data rows and then on the results. It provides instructions on how to run queries on file data and then on the results, and includes a list of "Other Query Options" such as "Download History & Recommendations" and "ICE: Update marked ids if files need it". A vertical line with a horizontal arrow points to the first item in a numbered list of steps.

Options

Quick | Process | Files | S1 Queries | S2 Queries | Opt Queries | ASC | Review / DEI | OK

General

These Options control the main flow of the process. You can select different query sets to run on the data rows from the file, and then on the row results. If you haven't selected a query set here, then it doesn't run. If you have selected a query set here, then it runs just the queries you have selected from the set, and the set runs in the way and order signified by this page.

First, run the following queries On The File Data to find as many matches as possible.

- Run Step 1 Queries. Call any results 'Set 1'
- Run Step 2 Queries. Put any results in 'Set 2'
- Run Optional Queries. Put the results in 'Set 2'

Other Query Options:

- Download History & Recommendations.
- ICE: Update marked ids if files need it.
- Force ICE : Always mark and need.
- Add pairs implicit in row matches.

After File Rows have been searched, repeat the following queries to find more matches, and id pairs, starting with the matches from above.

- Use Step 2 Queries in this process.
- Use Optional Queries in this process.

Check for all pairs and matches in this order upto (and including) this step:

- 1. Use just the results in 'Set 1' to find more results. Put them in 'Set 2'
- 2. Next use results from 'Set 2' to find more results. Call them 'Set 3'.
- 3. Next use the results you just got, and look for their matches. ('Set 4')
- 4. Last use the results from that and look for matches one more time. ('Set 5')

Files

Options

Quick | Process | **Files** | S1 Queries | S2 Queries | Opt Queries | ASC | Review / DEI | **OK**

Files

These options control what files you get, and the details of those files.

- Generate -mas.txt File, with records found by step 1 only. (very close matches.)
- Generate -pai.txt File, which contains unique possible-match ID pairs.
 - Add every pair that is indirectly connected by ids returned in the search.
- Generate -rec.txt File, which contains all the records found relating to each input row.
- Generate -not.txt File, filled with the rows from the input file which have zero matches.
- Generate -med.txt File, which follows the MEDD file specification.
- Generate -asc.txt File with the options on the ASC Fields tab.
- Generate -recall.txt File, which follows the Recall file specification.
- Generate -rrr.txt File, which follows the Recall file specification.
- Generate DEI Files, which follow the DEI file specification.
- Output only Recall and DEI files when reviewing records.

Include Rows from the input file in the -rec.txt file in the following cases:

- For rows with zero matches
- For rows with matches
- Place "[X]MATCHES.FOR " before the row.

SQL log: (choose one)

No SQL log. Normal SQL log. Detailed log with results. (a big file)

Generate -log.txt File, which is the Process Log. Ora Trace (select only when instructed to)

Default File Options

Step 1 Queries

Options

Quick | Process | Files | S1 Queries | S2 Queries | Opt Queries | ASC | Review / DEI | OK

Step 1 Queries

These queries are close matches and are applied on data rows only to find "close" matches. They run only when the process selects step 1 on the General Tab.

- Q1.a. Last, First, DOB ,Gender
- Q1.1: Medical Record #, DOB
- Q1.2: Medicaid
- Q1.3a: CIR id
- Q1.3b: CIR id in column 'cirb'
- Q1.4: Telephone Number

- Q1. i After Step 1 Queries, import pairs from file if both ids are valid.

Defaults

Select All

Unselect All

While searching with these queries as well as S2 and Opt queries using row data from the input file, allow only upto this number of matches per row. (0 means no limit.)

Limit:

If a box is checked then all the fields will be matched using the available fields when there are enough available fields.

Step 2 Queries

Options

Quick | Process | Files | S1 Queries | S2 Queries | Opt Queries | ASC | Review / DEI | OK

Stage 2 Queries

These queries are partial matches. The General Tab can apply these on file data and result rows.

<input checked="" type="checkbox"/> a0. Med Rec# + DOB	<input type="checkbox"/> d0. Last,First,DOB,gender
<input checked="" type="checkbox"/> a1. Med Rec# + (Month + Year of DOB)	<input checked="" type="checkbox"/> d1. Last,First,DOB
<input checked="" type="checkbox"/> a2. Med Rec# + (Date + Year of DOB)	<input checked="" type="checkbox"/> d2. Last,First,(Month + Year of DOB)
<input checked="" type="checkbox"/> a3. Med Rec# + (Date + Month of DOB)	<input checked="" type="checkbox"/> d3. Last,First,(Date + Year of DOB)
<input checked="" type="checkbox"/> a4. Med Rec# + First Name	<input checked="" type="checkbox"/> d4. Last,First,(Date + Month of DOB)
<input checked="" type="checkbox"/> c. Medicaid Number	
<input checked="" type="checkbox"/> c2. Medicaid Number + Med Rec#	
<input checked="" type="checkbox"/> c3. Medicaid Number + DOB	
<input checked="" type="checkbox"/> e. Last, First, Gender + (Month + year of DOB)	
<input checked="" type="checkbox"/> f. Mom's Maiden Name,First,DOB,Gender	
<input checked="" type="checkbox"/> g. Mom's Maiden Name,Mom's DOB, DOB, No mult. birth	
<input checked="" type="checkbox"/> h. Telephone, First, Gender	
<input checked="" type="checkbox"/> i. Fname, DOB, Gender, MomName = Orig Lastname	
<input checked="" type="checkbox"/> j. Fname, DOB, Gender, Lastname = Orig MomName	
<input checked="" type="checkbox"/> k Last, 1st 3 chars of First, DOB, Gender, No mult. birth	
<input checked="" type="checkbox"/> l. First, 1st 5 chars of Last, DOB, Gender	
<input type="checkbox"/> m. Last, DOB, Gender, No mult. birth	

Defaults

Select All

Unselect All

While searching with these queries using ids, limit results per id: (0 means no limit.)

each a	<input type="text" value="50"/>
each c	<input type="text" value="50"/>
d-g i-m	<input type="text" value="200"/>
for h	<input type="text" value="50"/>

If a box is checked then all the fields will be matched using the available fields when there are enough available fields.

Optional Queries

Options

Quick | Process | Files | S1 Queries | S2 Queries | **Opt Queries** | ASC | Review / DEI | OK

Optional Queries

These optional queries match slightly similar rows. They can be used to match file rows, and the match the data rows that earlier matches have returned. They run only when selected on the General Tab

o1. Medical Record # only

o2. First, Gender, DOB

o3. First, Gender, (Month + Year of DOB)

o4. First, Gender, (Date + Year of DOB)

o5. Last, Gender, DOB

o6. Last, Gender, (Month + Year of DOB)

o7. Last, Gender, (Date + Year of DOB)

o8. First 5 Letters of Last, DOB, Gender

o9. First 3 Letters of First, DOB, Gender

o10. Last, DOB

o11. First, DOB

o12. Gender, DOB, (original lastname = mom's maiden name OR original mom's name = lastname)

o13. Telephone

Limit common placeholder Firstnames

Defaults

Select All

Unselect All

While searching with these queries using ids, limit results per id: (0 means no limit.)

o1:

o2 - o12:

o13:

If a box is checked then all the fields will be matched using the available fields when there are enough available fields.

ASCII Fields & Options

Options

Quick | Process | Files | S1 Queries | S2 Queries | Opt Queries | **ASC** | Review / DEI | **OK**

ASC Fields and Options

If the asc file is selected to generate on the File tab, then write the file with these fields and options.

<input checked="" type="checkbox"/> Patient ID	<input checked="" type="checkbox"/> P.Key 3	<input checked="" type="checkbox"/> DTP 4	<input type="button" value="Defaults"/>
<input checked="" type="checkbox"/> Facility Code	<input checked="" type="checkbox"/> P.Key T 4	<input checked="" type="checkbox"/> HIB 1	
<input checked="" type="checkbox"/> Last Name	<input checked="" type="checkbox"/> P.Key 4	<input checked="" type="checkbox"/> HIB 2	
<input checked="" type="checkbox"/> First Name	<input checked="" type="checkbox"/> P.Key T 5	<input checked="" type="checkbox"/> HIB 3	
<input checked="" type="checkbox"/> DOB	<input checked="" type="checkbox"/> P.Key 5	<input checked="" type="checkbox"/> HIB 4	<input type="button" value="Select All"/>
<input checked="" type="checkbox"/> Gender	<input checked="" type="checkbox"/> P.Key T 6	<input checked="" type="checkbox"/> Polio 1	<input type="button" value="Unselect All"/>
<input checked="" type="checkbox"/> Mom M Name	<input checked="" type="checkbox"/> P.Key 6	<input checked="" type="checkbox"/> Polio 2	
<input checked="" type="checkbox"/> Mom DOB	<input checked="" type="checkbox"/> Father LN	<input checked="" type="checkbox"/> Polio 3	
<input checked="" type="checkbox"/> House num	<input checked="" type="checkbox"/> Father FN	<input checked="" type="checkbox"/> Polio 4	
<input checked="" type="checkbox"/> Street	<input checked="" type="checkbox"/> Mother LN	<input checked="" type="checkbox"/> MMR 1	<input checked="" type="checkbox"/> Other 1
<input checked="" type="checkbox"/> ZIP Code	<input checked="" type="checkbox"/> Mother FN	<input checked="" type="checkbox"/> MMR 2	
<input checked="" type="checkbox"/> City	<input checked="" type="checkbox"/> Gaurdian LN	<input checked="" type="checkbox"/> MMR 3	
<input checked="" type="checkbox"/> State	<input checked="" type="checkbox"/> Gaurdian FN	<input checked="" type="checkbox"/> MMR 4	
<input checked="" type="checkbox"/> Zip Code	<input checked="" type="checkbox"/> HepB 1	<input checked="" type="checkbox"/> Var. 1	<input checked="" type="checkbox"/> Other 2
<input checked="" type="checkbox"/> Telephone	<input checked="" type="checkbox"/> HepB 2	<input checked="" type="checkbox"/> Var. 2	<input checked="" type="checkbox"/> Other 3
<input checked="" type="checkbox"/> P.Key T 1	<input checked="" type="checkbox"/> HepB 3	<input checked="" type="checkbox"/> Var. 3	<input checked="" type="checkbox"/> Other 4
<input checked="" type="checkbox"/> P.Key 1	<input checked="" type="checkbox"/> HepB 4	<input checked="" type="checkbox"/> Var. 4	
<input checked="" type="checkbox"/> P.Key T 2	<input checked="" type="checkbox"/> DTP 1		
<input checked="" type="checkbox"/> P.Key 2	<input checked="" type="checkbox"/> DTP 2		
<input checked="" type="checkbox"/> P.Key T 3	<input checked="" type="checkbox"/> DTP 3		

Field Delimiter:

Review

Options

Quick | Process | Files | S1 Queries | S2 Queries | Opt Queries | ASC | Review / DEI | OK

Record Review Options

To review a record set, specify the ss file (ends with "ss.txt") on the main page and hit "Review"

Autosave will export the files from review just like selecting "Save" as often as you specify. Enter 0 to prevent autosave.

Autosave time in minutes =

DEI Field Delimiter

Mark DEI Recommendations "REC" after the series name

Record Review

Recall Record Review

File Options

Orig. Row:

PAPADOUKA VIKKI 01/01/1999 F

Possible Matches:

1	X	248513512 N PAPADOUKAS VASSILIKI 01/01/1999 F ALEXANDRA 01/01/1930 123 VP02135A DB ADDRESS: NY
2	X	248702951 N PAPAD VASSILIKI 01/01/1999 F DB ADDRESS:
3		248702387 N PAPADOUKA VASSILIKI 01/13/1999 F DB ADDRESS:
4		248702736 N PAPADOUKA BABY GIRL 01/01/1999 F DB ADDRESS:
5	X	248702512 N PAPADOUKA VIKKI 01/01/1998 F 123 DB ADDRESS:
6		248702624 N MATTHEWS VASSILIKI 01/01/1999 F PAPADOUKA 01/01/1930 123 VP02135A DB ADDRESS:
7		248703063 N PAP VASSILIKI 01/01/1999 F DB ADDRESS:
8		248791375 N SPENCER SHIRLEY 02/09/1999 F 123 DB ADDRESS: NY

1 row

next mult row

next mult sel

next no sel

next long

1 / 1

Changed

Conclusion

- ♦ **Smart Search** is a powerful query tool which is a vital component of our de-duplication and data quality efforts, and our efforts to utilize Immunization Registry data