Patient Data Matching Software: A Buyer's Guide for the Budget Conscious

Prepared for:

CALIFORNIA HEALTHCARE FOUNDATION

Prepared by:

Sujansky & Associates, LLC

Authors:

LeRoy Jones, M.S.

Walter Sujansky, M.D., Ph.D.

ihealthreports

August 2004

About the Author

LeRoy Jones, M.S. is a principal at Safe Outsourcing, an information-technology services company. He leads the Federal Health Architecture program within the Office of the National Coordinator for Health Information Technology.

Walter Sujansky, M.D., Ph.D. is president and senior consultant at Sujansky & Associates, LLC, a consulting firm specializing in health care informatics and software development (www.sujansky.com).

About the Foundation

The **California HealthCare Foundation**, based in Oakland, is an independent philanthropy committed to improving California's health care delivery and financing systems. Formed in 1996, our goal is to ensure that all Californians have access to affordable, quality health care. For more information, visit us online at **www.chcf.org**.

This report was produced under the direction of CHCF's Chronic Disease Care program with support from the Health Information Technology group. The research was conducted to support activities of the California Clinical Data Project: Setting Standards. Visit **www.chcf.org/ setting standards** for more information.

The iHealth Reports series focuses on the effective adoption of IT in health care by analyzing the marketplace, inspiring innovation, and providing practical information on emerging technology trends.

ISBN 1-932064-74-5

Contents

4 Executive Summary

6 I. Introduction

7 II. Audience and Background

9 III. Identification of Candidate Patient-Matching Products

11 IV. Readiness Checklist

13 V. Evaluation Procedure

15 VI. Evaluation Results

25 VII. Summary and Recommendations

27 Appendix: Test Data Used for Product Evaluations

Executive Summary

THE CALIFORNIA HEALTHCARE FOUNDATION asked Sujansky & Associates to investigate cost-effective patient-matching software that can assist small and medium-size provider organizations in the development of disease registries and clinical data repositories. These tools apply probabilistic and fuzzy-matching techniques to link records in disparate data files to the correct patient identities. (As opposed to methods that rely on exact matches, fuzzy-matching techniques can identify "near matches" and use them to determine whether two records are likely to represent the same person.) The patient-matching task is an important and challenging step in the creation of integrated clinical databases used for quality-measurement and quality-improvement purposes.

Many provider organizations have locally developed programs that use relatively basic and ad hoc matching algorithms. The rates of success of these programs may be lower than desired. Cost-effective commercial products that apply more sophisticated, state-of-the-art methods could improve success rates, as well as provide a means to perform patient matching for organizations that have not yet developed their own programs.

The research identified candidate products that are commercially available and assessed these products with respect to provider organizations' specific requirements for patient matching. These requirements were ascertained through interviews with five provider organizations that currently perform patient matching in the course of developing clinical data repositories. The size of these organizations ranges from 7,400 to 340,000 covered lives.

The authors identified the following high-level requirements: ease of use; availability on a desktop platform; application of advanced matching techniques; ability to integrate into existing patient-matching workflows; and a total cost of ownership not exceeding $50,000. Based on these requirements, they identified four candidate products and assessed each of them in detail (including hands-on matching of test data sets):

- LinkageWiz (LinkageWiz Software)
- SureMatch (DQ Global)
- DataSet V Suite (Intercon Systems)
- DeDupe4Excel (DQ Global)

This guide documents the assessment. It includes: 1) a qualitative description of each product's capabilities with respect to 20 relevant features; 2) a head-to-head quantitative scoring of the products with respect to the same features; and 3) an inventory of each product's capabilities with respect to the provider organizations' requirements.

The conclusions indicate that three of the four products are good candidates for the patient-matching needs of most provider organizations. These products are LinkageWiz, SureMatch, and DataSet V Suite. Because these products vary somewhat in their ease of use, ability to be customized, and costs of ownership (the cost of these products ranges from $350 to $11,000), the authors encourage prospective buyers to try out each of the tools via the demonstration copies available from the vendors.

The fourth product, DeDupe4Excel, may not be suitable for most provider organizations because it can process a maximum of 64,000 records at one time. Although this limitation may be prohibitive for medium to large organizations, very small organizations may find it acceptable. Therefore a description of DeDupe4Excel is included.

I. Introduction

Effective matching requires the use of probabilistic and fuzzy-matching techniques.

INFORMATION TECHNOLOGY AND THE USE OF electronic clinical data can facilitate quality-improvement efforts in health care. Because electronic medical record (EMR) systems have not yet become a mainstream technology, many health care organizations have developed useful clinical databases by integrating electronic data from a variety of existing sources, rather than capturing the data in EMRs. Existing sources include ambulatory and hospital encounter data, pharmacy claims records, and laboratory test reports. To successfully integrate these data, health care organizations must correctly match records from the data sets they receive to their patient roster. This patient-matching task is critical because the establishment of accurate and complete patient profiles is a prerequisite for useful data analysis.

In practice, matching data from multiple, independent data sources is challenging and error-prone, even for organizations with information technology experience and resources. Because no standard patient identifier exists within the private health care system, clinical data must be matched based on multiple, imprecise data elements, such as name, date of birth, health plan ID, and medical record number. Values of these identifying attributes may be shared by multiple patients, represented inconsistently across data sources, and subject to change over time. To accommodate these inconsistencies and variances, effective matching requires the use of probabilistic and fuzzy-matching techniques.

This report describes several moderately priced, commercially available software products that can assist organizations in the patient-matching task. These tools all apply advanced patient-matching techniques and are relatively easy to use. For health care organizations that have not yet developed their own patient-matching tools or who are dissatisfied with the effectiveness of their current tools, these products may serve as useful starting points, alternatives, or supplements for their patient-matching processes.

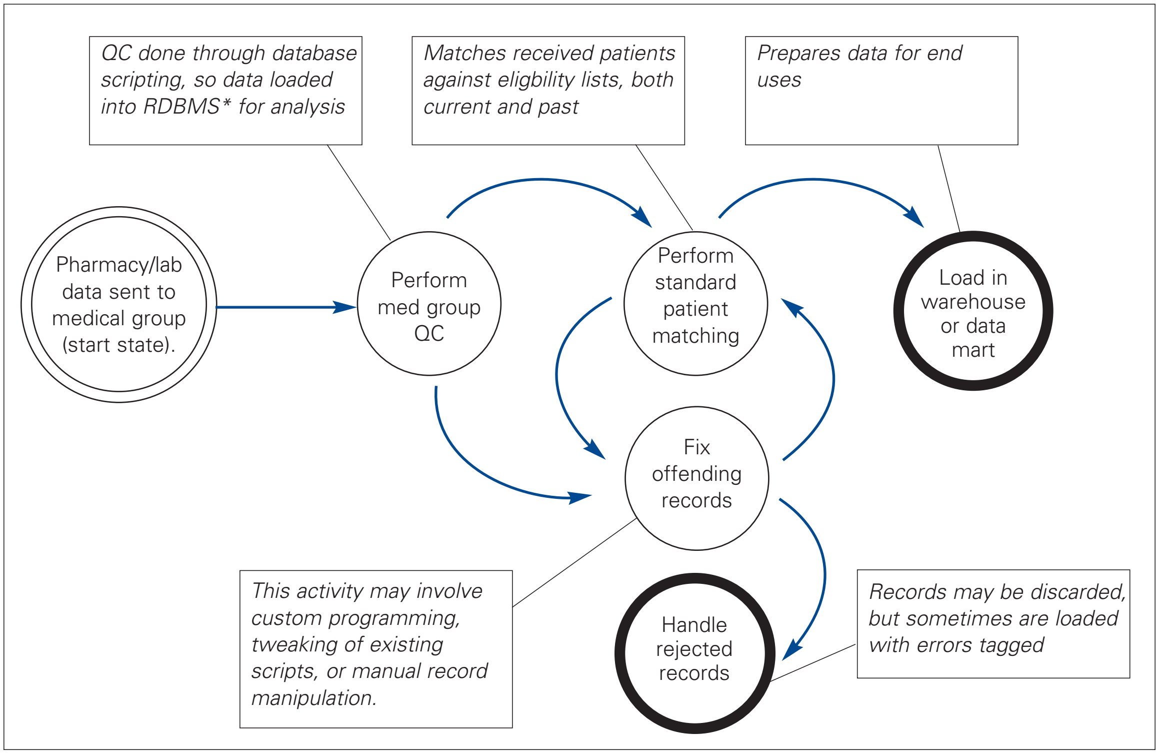
II. Audience and Background

Many health care organizations face the challenges of patient-matching, including health plans, individual hospitals, integrated delivery networks, and regional health care systems. This report, however, is intended to address the needs of a specific audience pursuing a specific goal: *ambulatory provider organizations seeking to develop clinical data repositories for purposes of quality measurement and quality improvement.*

This audience consists of independent practice associations (IPAs), community clinics medical groups, or consortia thereof. Many of these organizations have developed or plan to develop clinical data repositories to demonstrate and/or improve their quality of care. These repositories draw clinical data from existing sources, such as encounters, pharmacy claims, and lab results—then match these data to patients listed in insurance eligibility files. The resulting databases provide patient-specific clinical profiles that support data analysis and reporting.

The following diagram shows the patient-matching process as it is typically performed at these types of organizations (the process begins with circle at left):

Figure 1. Typical Patient-Matching Data Flow



QC done through database scripting, so data loaded into RDBMS* for analysis
Matches received patients against eligbility lists, both current and past
Prepares data for end uses
Pharmacy/lab data sent to medical group (start state).
Perform med group QC
Perform standard patient matching
Load in warehouse or data mart
Fix offending records
This activity may involve custom programming, tweaking of existing scripts, or manual record manipulation.
Handle rejected records
Records may be discarded, but sometimes are loaded with errors tagged

* Relational database management system, e.g., Microsoft, Access, Oracle.

The authors ascertained the requirements of patient-matching tools for this audience by interviewing five California organizations that are currently integrating patient data into clinical data repositories. Brown & Toland Medical Group, Greater Newport Physicians, Hill Physicians, Humboldt Del Norte Physician Group, and Intelligent Healthcare, LLC. These all use eligibility files containing lists of the patients whose care they manage. They match the laboratory, pharmacy, and claims data they receive against these eligibility files to associate the data with the patients known to them.

To perform matching, the organizations typically load the data into a relational database, where scripts or programs are executed to identify the matches. The identifying data elements typically used are: first name, last name, date of birth, and health plan member ID. Each of the five organizations uses matching algorithms that it has designed and programmed; none uses a commercial product or standard probabilistic matching techniques. The algorithms vary from simple "exact match" schemes to more advanced algorithms involving weighting of fields based on historical matches. The results of these processes, as self-reported, vary from a match rate as high as 98 percent to much lower rates. The degree of manual review and editing varies.

Clinical data that cannot be matched against any records in the eligibility files usually remain "orphaned" in the database unless and until a match is made. Therefore they are not available for analysis or reporting.

There is no standard computing environment across the organizations. The computer systems involved in patient matching range from mainframes to desktops, and the tools vary considerably. However, a relational database is commonly used as a staging area for patient matching and a repository of matching results, and the Microsoft Windows desktop environment is the tool of choice for the data analysts who perform this work.

III. Identification of Candidate Patient-Matching Products

BASED ON ANALYSIS OF PROVIDER ORGANIZATIONS' requirements, the authors developed the following criteria for screening stand-alone commercial patient-matching tools:

- **Cost under $50,000.** This is to accommodate the largest proportion of purchasers, although it is recognized that a minority of provider organizations may be able to afford more expensive tools.
- **Use of advanced matching algorithms.** All of the tools reviewed use some combination of phonetic, orthographic (the study of spelling and its variations), probabilistic, and/or fuzzy-matching techniques. The goal was to find a set of tools that provides more sophisticated matching than the typical home-grown solutions, which often are based only on comparison of substrings or concatenation of substrings across a small number of data fields (for example, first name+last name+date of birth).
- **Ability to match based on parameters other than names and addresses.** Many of the lowest-priced tools are designed to de-duplicate mailing lists for mass-marketing purposes, and are sometimes limited to processing mail-label fields. This is not sufficient for medical groups that have additional parameters at their disposal (such as date of birth, medical record numbers, etc.), and which need more accurate matching than the typical mass-marketing application.
- **Ability to export findings into a usable form for subsequent processing**. The results of the matching process must be in a form that may be readily integrated back into the data-integration workflow. Some tools produce reports that are readable by people, but not structured for subsequent processing by scripts or database-import tools.
- **Ability to be installed and run on a Windows-based PC.** Interviews with data analysts performing the matching function indicated that the task is often done by a small number of individuals working independently, using whatever tools are at their disposal. These people often incorporate their personal workstations into the workflow, and most use Microsoft Windows.
- **Availability for hands-on evaluation.** There is no widely accepted method to evaluate how well a tool of this kind performs, but it was essential to interact with each tool directly to assess its effectiveness and ease of use.

Using these criteria, the authors identified suitable tools in the fall of 2003 by searching the Internet and publicly available registries of health information software, such as the Healthcare IT Yellow Pages (www.health-infosys-dir.com/yp_hc.asp). This search entailed review of information posted to the Internet, as well as follow-up questions to vendor representatives. The search was conducted using the Google and AltaVista search engines. Search terms included:

- "patient matching"
- "identity matching" AND "healthcare"
- "record linking" AND "healthcare"
- "master patient index"
- "master person index"

This search identified many tools that met the cost criterion, but did not provide sufficient functionality. For example, many data-cleansing tools remove duplicates from mailing lists and customer databases for mass-marketing purposes (an example is WinPure's ListCleaner product, available for $249). Most of these tools, however, apply simplistic matching methods, use name and address fields only, and offer limited configurability. Such limitations make these tools inappropriate for the patient-matching task, in which additional data elements are available and more precision is required.

Also identified were several tools that provide advanced matching methods and extensive configurability, but entail licensing costs well above the $50,000 threshold. These software products, such as Initiate Systems' Identity Hub, provide impressive functionality for building and maintaining enterprise master-person indexes. However, they require significant resource commitments for system integration and configuration, as well as software licensing fees often in the multiple six figures. These products are used by many large health care organizations, and with more extensive patient-matching needs and more financial resources may wish to explore such products.

Based on the criteria developed for small and medium-size provider organizations and the search of publicly available resources, the authors identified and evaluated the following commercial products:

- LinkageWiz (LinkageWiz Software)
- SureMatch (DQ Global)
- DataSet V Suite (Intercon Systems)
- DeDupe4Excel (DQ Global)

In general, all of these tools involve a similar sequence of steps:

- Import the data.
- Massage the data to facilitate a field-by-field comparison of records.
- Specify match weights for the relevant demographic and other fields.
- Run a number of matching algorithms against the data and compute matching scores that indicate the likelihood of a record-pair match.
- Display the actual and possible matches for manual (clerical) review and editing.
- Export the set of matched records for further processing and data integration.

All of the identified tools assume that users have modest familiarity with data processing and operate through a graphical user interface. Three of the four products offer both a de-duplicate mode (duplicate entries in a single file are detected and removed) or a matching mode (matching entries in two distinct files are detected and reported, with one file acting as the master and the other as the reference file). The fourth product, DeDupe4Excel, operates in only the de-duplicate mode, but the user can simulate the matching mode by concatenating the master and reference files.

IV. Readiness Checklist

THE TOOLS DESCRIBED IN THIS GUIDE ARE relatively flexible and easy to use. However, there are several prerequisites to the effective application of these tools to the patient-matching task:

- **Basic understanding of probabilistic and fuzzy-matching techniques.** Although the reviewed tools are relatively user friendly, most of them require some configuration of matching weights and desired data transformations in order to optimize the matching process (default weights and transformations are also provided in most cases). The documentation provided with the products assumes some basic understanding of probabilistic patient-matching principles and terminology (e.g., familiarity with terms such as "exclude lists" and "blocking variables").
- **Availability of data in a tabular format.** The data to be matched must be available in a tabular format that can be processed or imported by the tools. All of the tools accommodate data in a Microsoft Access database or Microsoft Excel spreadsheet. Some also handle data in delimited text files or in any relational database that can be accessed via ODBC (open database connectivity).
- **Existence of a master patient file with unique identifiers.** The tools assume that one of the files to be matched is a master file, containing the organization's list of known patients. Typically, this file has been preprocessed so that it contains no duplicate identities (most of the tools provide a de-duplication capability for single files).
- **Familiarity with the values of demographic fields in the data**. To configure the tools appropriately, the user must have some familiarity with the specific values of data elements that are available for matching in the master file and the input (reference) file. For example, to configure the tool to be most accurate and efficient, the user should know which fields are more and less accurate, which have common synonyms, which are sometimes omitted, and which are more specific to individuals.

- **Ability to post-process the output of the matching tools.** Most of the tools described in this report cannot directly load the matched records into a clinical data repository. All of the tools generate a matching report that can be exported to a file (usually an Excel spreadsheet or text file). These files contain field values from the records that were determined to match. To integrate the clinical records into a data repository, it is necessary to post-process these export files. In most cases, this can be done using standard database-import and querying tools, such as SQL.

V. Evaluation Procedure

There were three steps in the evaluation of the tools.

1. Qualitative assessment. Each tool was examined with respect to the features required by provider organizations for patient matching. The information from this assessment is documented in Table 2, a side-by-side comparison of the products. Table 1, below, shows the requirements that were used for this part of the evaluation.

Table 1. Requirements for Patient Matching

Input Specifications	
■ Ease of import	How well does the tool allow for the import of data?
■ Flexibility of import formats	What formats are supported as valid sources?
■ Health care-specific accommodation	Are there any features built into the tools specifically to support health care applications?
■ Amount of data supported	What volume of data can the tool handle?
Matching Capability	
■ Number of sources allowed	How many sources may be compared simultaneously?
■ Sophistication of matching algorithms	What are the standard methods used by the matching process?
■ Optimization for common field types	What predefined data types are recognized by the tool?
■ Degree of user configurability	What options are available for the user to tailor the matching process?
■ Flexible definition of "sameness"	Can the user influence the relative contribution of each data field to the determination of a match?
■ Speed	What is the throughput rate?
■ Matching against test data	How does the tool perform versus the human (the gold standard) matching process?
Post-match Processing	
■ Clarity of presentation of matching results	How understandable and useful are the matching reports?
■ Flexibility in configuring end result	What options exist for the user to define the content or form of the output?
■ Export target formats	In which formats can results be exported?
Supported Platforms and System Requirements	What platform or environment is supported by the application?
Extensibility	Are there any hooks to extend the software functionality programmatically or to integrate it with other software?
Documentation and Support	How helpful and convenient is the documentation provided with the product? How available are technical support and consulting services?
Strengths	What are the best features of the software?
Weaknesses	What are the limitations or worst features of the software?
Pricing	How much does the software cost? Is it a good value?

2. Quantitative scaled assessment. The following scale was used to quantitatively assess the tools with respect to the same features:

1	2	3	4	5
Unacceptable	Inadequate	Adequate	Good	Excellent

This assessment measured how well a given tool supports the needs of the provider organizations for patient matching. Individual scores for each relevant feature, as well as an aggregate score across all features, are reported in Table 3.

To perform consistent evaluations, a sample pair of data files was compiled from two separate sources of phonebook data (see Appendix). These sources agreed with one other to a certain degree, but differed in field formatting, spelling, data transposition, and other content variations. Also, synthetic birth dates were added to the records to allow records to have similar but not identical birth dates. The names were chosen from a popular Asian name, Ngu, which acts as a prefix for similar Asian names. These choices in developing the test data allowed researchers to exercise the tools against difficult matching decisions.

One file of 50 records served as the master file, and a second file of 40 records served as the input (reference) file. The composition of the test files allowed researchers to test the ability of the tools to accommodate phonetic and orthographic variations in data values, as well as missing data. Although the test database was very useful in exercising each tool's features, it is not large enough to accurately measure the speed or to assess the absolute or relative accuracy of the matching algorithms in the general case. The speed of the tools against this test data set are noted in the evaluation only to identify egregious performance problems.

3. Assessment of the products' capabilities against the requirements of provider organizations. This phase determined which of the requirements were fully or partially met by each tool. The results of this assessment appear in Table 4.

VI. Evaluation Results

1. Qualitative Assessment. The following table provides a comparison of the four products as measured against the requirements for patient matching.

Table 2. Side-by-Side Product Comparisons

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Input Specification				
Ease of import	Importing data is very straightforward from a number of formats; fields are detected, and then user maps them to defined types.	Importing is accomplished by a wizard that walks user through each option step-by-step per field, or self-directed through tab interface. Data is required to have a single-field primary key for record uniqueness, which may require data to be added to if not present. Wizard can be confusing because some options are for advanced users and are rarely relevant. Self-directed panel is friendlier. Evaluated version had small bug in wizard.	Data must reside in Excel sheet with column labels as first row; no explicit import feature other than those of Excel.	Importing is fairly involved, as the tool creates a master data store from user involvement in a wizard-like process that requires both specifying and saving of several files (atypical for this type of tool).
Flexibility of import formats	MS Access; Dbase III, IV, V; Paradox 3.x - 5.x; MS Excel 3.x - 5.x, 97; text (.csv, .txt, .dat).	Excel 3.x - 5.x, 97, 2000; Lotus 1 2 3 (wk3); MS Access 2.x, 95, 97, 2000; Dbase III, IV, V; Foxpro 2.0, 2.5, 2.6; Paradox 3.x - 5.x; text (.csv); text (fixed length); Foxpro 3.0 (only via 32-bit ODBC); Btrieve (only via 32-bit ODBC); ODBC (requires 32-bit drivers.	Same as MS Excel.	MS Access; text; MS Excel; ODBC for master data store; limited to MS Access only for matching source.

Table 2. Side-by-Side Product Comparisons (cont.)

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Health care-specific accommodation	Medicare number, and diagnosis codes as predefined types, but apply to Australian conventions for same.	None.	None.	None.
Amount of data supported	Physical limit of 6-10 million records, depending on record width. MS Access-based tool with 2GB database/table size limit.	No measured limit.	Number of records limited to row limit of MS Excel (c. 65,536 rows). This may be inadequate for all but the smallest of eligibility lists.	No measured limit. Purportedly handling terabytes of data in an MS Access database with a data compression add-on, but MS Access has physical limit of 2GB database/table size.
Matching capability				
Number of sources allowed	One source for de-duplicating, or two sources for matching.	One source for de-duplicating, or two sources for matching.	All data must be contained in a single sheet (one source).	One source for de-duplicating, or two sources for matching.
Sophistication of matching algorithms	Probabilistic record linkage algorithms. Phonetic name matching using NYSIIS and SOUNDEX algorithms; string comparator functions; automatic nickname translation; user-definable linkage variables and linkage weights; optionally use diagnostic fields as linkage fields (cancer registry and hospital customers); value-specific linkage weights (e.g., the family name 'Smith' receives a lower weight than 'Fellegi.'	Uses phonetic matching, and fuzzy non-phonetic matching. Does elaboration on fields, and reduction to short forms in order to determine fuzzy inclusion.	Uses phonetic algorithms (sound-alike), common abbreviation expansion, extraneous character removal, and fuzzy comparisons to match fields.	Tool allows for comparisons through factoring of data fields into many categories and component parts, thereby standardizing the input for fuzzy selection. Has a unique specialty comparison, the geographical match, which will use spatial proximity as a criterion for matching, if desired, on geographic fields such as postal codes and addresses.
Optimization for common field types	Includes various forms of names (nickname, given name, first, middle, etc.); dates of birth and death; gender; address; postal code; Medicare number; diagnoses; user-defined types.	Includes names, addresses, organizations, telephone numbers, email, and postal codes.	Includes dates, addresses, names (suitable for first names), postal codes, dates, phone numbers, company names, and titles. Does simple normalization on those predefined types.	Includes dates, addresses, justified numbers, phone numbers, zip codes, email, cities, and states.

Table 2. Side-by-Side Product Comparisons (cont.)

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Degree of user configurability	User may specify what fields to consider, how they map from source ("master") file to reference file, what predefined type a field corresponds to. User may define the sequencing of fields used in identifying matches in the multiple-pass system. The tool "prepares" the databases for matching, which includes creating additional fields to hold expansions and derivations of the data, and the user may edit any aspect of the database as a lower-level way to influence matching. Frequencies of field values can be examined to account for default values. Thresholds for match scores are handled after matching through analysis of the data.	User may specify what fields to consider, what predefined type a field corresponds to, and the matching threshold. Includes options to specify fields to have fuzzy, phonetic, or exact matching rules applied where no predefined type exists, or may not yield the desired results. User can redefine how predefined types are processed in terms of the matching algorithms that apply to them; 11 categories of transformations (lists of synonyms) exist for normalization/elaboration, such as addressing, qualifications, salutations, events, first names, countries, job titles, first names numbering, etc.	User may specify what fields to consider, what predefined type a field corresponds to, and the matching threshold.	User may specify what fields to consider, how they map from source ("master") to reference, what predefined type a field corresponds to. User may define the sequencing of fields used in identifying matches in the multiple-pass system. User may specify synonym lists per field, exclusion lists on field values, and delimiter filters for extraneous characters.
Flexible definition of "sameness"	User can set individual weights given to predefined types and user-defined types; for predefined types, weights may be adjusted for the various matching rules applicable to them.	User cannot influence how matches are determined through weighting.	User cannot influence how matches are determined through weighting.	User can set individual weights given to predefined types and user-defined types for both their confirmatory and exclusionary contribution to matching.

Table 2. Side-by-Side Product Comparisons (cont.)

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Speed	Approximately 22 records per second. Tool allows a small subset to be processed initially, to ascertain time required. Allows specification of blocking variables to improve efficiency.	Approximately 25 records per second, with seven fields included in the matching process.	Approximately 25 records per second, with seven fields included in the matching process.	Approximately 16 records per second, with seven fields included in the matching process. Tool allows user to suspend and resume the matching process. Allows specification of blocking variables to improve efficiency.
Matching against test data[1]	All of the tools performed competently, discovering non-obvious matches based on the full set of available matching fields. In matching the test data, all of the tools significantly outperformed a rudimentary deterministic matching algorithm that is known to be used by at least one provider organization.[2] Actual matching statistics, however, are not presented here because the test set used for this evaluation was very small and artificially generated (see Appendix). Results from this test set may not be representative of the data typically processed by provider organizations and could mislead the reader. Prospective buyers should try demonstration versions of several of the tools on their specific data sets.			
Post-match processing				
Clarity of presentation of matching results	Choice of static report showing number of linkages, or spreadsheet-like display allowing sorting, filtering, etc. Frequency graph of matching scores available to determine proper threshold for true matches. Information not presented in a very straightforward manner; difficult to interpret at a glance.	Very understandable report of duplicates, using color shading to distinguish groups, and a bisected table to distinguish between the master record, and duplicates. Records can be edited to manually merge or change data. Summary view shows the results of the matching process overall, including the elapsed time, the count and percentage of duplicates, and the number of records from each source.	Separate spreadsheets are generated listing the records considered duplicates (matched), the matched groupings with the master record and its proposed duplicates, and the de-duplicated list. Clearly organized and labeled with matching scores.	Tool provides a "matching-review console" that allows user to see which record pairs matched with high, medium, or low probability. The console provides excellent functionality to review and edit the matching results.

[1] See Appendix

[2] This method sought an exact match between a derived field consisting of the concatenation of the first three characters of the last name, the first two characters of the first name, and the six-digit date of birth.

Table 2. Side-by-Side Product Comparisons (cont.)

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Flexibility in configuring end result	User can control what is reported (matches) as well as what is actually linked in the database. Several thresholds for matching score may be set by user, including: save a link, assign a link to a group, and report a link.	Several options for the final form including key lists that allow only the primary key values to be exported for the duplicates, masters, or both. The same options exist for the full records. Merging feature allows user to determine how the final records are populated, selectively pulling data from both/either the master and duplicate records.	No flexibility in tailoring output other than setting threshold for matching score.	Several thresholds may be set to partition results by matching score. Categories are: confirmed, high probability, low probability, above threshold, and unmatched.
Export target formats	Choose fields from database to export. Formats supported are .csv for ungrouped results, and .rtf for grouped report. Filters may be applied to fields to limit exported records.	All fields in record exported to a delimited text file, such as .csv, organized in various groupings of master and duplicate records.	Same as Excel.	Choose fields from master data store for export. Export specific matching results by categories. Export to Excel, Access, text, HTML. Export data mining reports by graphs and tables.
Supported platforms and system requirements				
	MS Windows OS on a PC-compatible workstation (Pentium 4 PC or greater recommended for larger files).	MS Windows OS on a PC-compatible workstation.	Add-in for Excel 97 and later (requires prior installation of Excel); MS Windows OS on a PC-compatible workstation.	MS Windows OS on a PC-compatible workstation.
Extensibility				
	No apparent programmatic extensibility, but technical services offered by vendor.	Corporate version is extensible via VB Script; purchase of a companion tool (SureData Toolkit) allows further extensibility, and creation of custom matching rules. SQL filters may be used to tailor the retrieval process.	Available within Excel, so as extensible as Excel.	No apparent programmatic extensibility, but technical services offered by vendor.

Table 2. Side-by-Side Product Comparisons (cont.)

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Documentation				
	A 113-page user manual and online help are provided. Excellent descriptions of the linking methodology, the product features, and tips to produce optimal linking.	Online help only. Explanation is minimal, and many product features are not described.	Online help only. Explanation is minimal, and many product features are not described	A 168-page user manual is provided. Comprehensive documentation, with screenshots and examples. Explanations of features are sometimes cryptic.
Support				
	4-10 hours of email support is included, depending on the version of the product purchased. Support personnel located in Australia. Additional support is available at extra cost: Block of 10 support hours: $550 Block of 20 support hours: $1,000 General contact: support@linkagewiz.com	Email and phone support provided, including during 30-day evaluation period. Support personnel located in U.K. General contact: support@dqglobal.com (011) 44-1329-227505	Email and phone support provided, including during 30-day evaluation period. Support personnel located in U.K. General contact: support@dqglobal.com (011) 44-1329-227505	Email and phone support provided, including during evaluation period. Additional consulting support may be purchased on hourly basis. Support personnel located in U.S. General contact: info@interconus.com 610-516-1625 or support representative: spaterson@interconus.com 610-516-1673
Strengths				
	Small Graphical User Interface (GUI) footprint; reasonable online help; good control over matching process.	Very clear reporting of results; understandable options for sufficiently configuring tool, balancing well the tool's ease of use and learning curve with the user's degree of control over the process.	Extremely easy to use (2 mouse clicks); very clear reporting of results; very inexpensive; takes advantage of widespread familiarity of Excel to lessen learning curve.	Very inexpensive for a stand-alone tool; user has a lot of control over the manner in which data are interpreted and matches are reviewed and edited.

Table 2. Side-by-Side Product Comparisons (cont.)

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Weaknesses				
	Slightly convoluted process for dealing with output; implied process for configuring tool not obvious via user interface.	Unnecessarily mixes different levels of abstraction in the user interface by having advanced options that require knowledge of SQL intermingled with mainstream functionality for non-technical user; requires primary key to be present in data where it could easily be generated; few export options.	Limited control over matching process; tied to Excel exclusively; rigid in file format.	Somewhat cluttered and unintuitive user interface; no online help; user manual is comprehensive, but does not explain functionality clearly in all cases.
Pricing				
	$3,495 for Enterprise Edition, unlimited records; $1,995 for up to 500,000 records.	$3,684 for Professional Edition; $7,377 for Enterprise Edition (2 sources allowed); $11,071 for Corporate Edition (VB scripting allowed); all pricing given is for single user, with discounts up to 30% for simultaneous purchase of additional users; unlimited records.	$349 per user, and each user must have a copy of MS Excel installed.	$750 per installation, unlimited users, unlimited records. This represents an excellent value, given the features of the tool.
Company information				
	LinkageWiz Software 24 Albert Street Payneham South Australia 5070 http://www.linkagewiz.com Tel: +61 41-1203-199 Fax: +61 8-8363-1861	DQ Global Ltd Cams Hall Cams Hill Fareham, Hampshire PO16 8AB United Kingdom http://www.dqglobal.com/ Tel: +44 (0)1329 227505 Fax: +44 (0)1329 227506 Note: Product available through on-shore resellers, such as: DQMax 9836 E. Baryte Pl. Tucson, AZ 85749-8168 Tel: 520-884-7778	DQ Global Ltd Cams Hall Cams Hill Fareham, Hampshire PO16 8AB United Kingdom http://www.dqglobal.com/ Tel: +44 (0)1329 227505 Fax: +44 (0)1329 227506	Intercon 790 Penllyn Pike, Suite 302 Blue Bell, PA 19422 (Corporate headquarters are in Jerusalem, Israel) http://www.ds-dataset.com/ Tel: 215-628-3700 Fax: 215-628-2754

2. Quantitative Assessment. The following table scores the four products in terms of their fulfillment of the requirements for patient matching.

Table 3. Scaled Scoring of Products

	LinkageWiz (LinkageWiz Software)	**SureMatch** (DQ Global)	**DeDupe4Excel** (DQ Global)	**DataSet V Suite** (Intercon Systems)
Input Specification				
Ease of import	4	3	4	3
Flexibility of import formats	4	5	2	3
Health care-specific accommodations	4	3	3	3
Amount of data supported	3	4	2	5
Matching capability				
Number of sources allowed	4	4	2	4
Sophistication of matching algorithms	4	4	4	4
Optimization for common field types	4	3	3	4
Degree of user configurability	4	3	2	5
Flexible definition of "sameness"	4	2	2	5
Speed	4	3	3	4
Matching accuracy against test data	3	3	3	3
Post-matching processing				
Clarity of presentation of matching results	3	3	3	5
Flexibility in configuring end result	3	3	2	5
Export capabilities	2	3	3	5
Supported platforms and system requirements	**4**	**4**	**3**	**4**
General				
Extensibility	3	4	4	3
Documentation and support	4	2	2	4
Product pricing	**4**	**3**	**5**	**5**
Product overall score	**65**	**59**	**52**	**74**
Product average score	**3.6**	**3.3**	**2.9**	**4.1**

3. Capability Assessment. The following table shows which of the requirements are fully or partially met by each tool.

Table 4. Features versus Patient-Matching Requirements

Requirement	LinkageWiz	SureMatch	DeDupe4Excel	DataSet V
The system shall provide a numerical score representing the degree of surety that the system judges any two presented identities to be the same.	Fulfills.	Fulfills.	Fulfills.	Fulfills.
The system shall use a standard definition for patient identity drawn from available characteristics of patients stored in the organizational database or available through data feeds.	Fulfills.	Fulfills.	Fulfills.	Fulfills.
The system shall be able to render a judgment on the degree of matching between two identities in the face of missing information from the patient's identity definitions.	Fulfills.	Fulfills.	Fulfills.	Fulfills.
The system shall be able to be tuned for increased accuracy based on information given it about known matches (accuracy defined as the percentage of time the system agrees with or convinces an established expert of patient matching).	System does not make use of historical data, but is tunable.	System does not make use of historical data, but is tunable.	System does not make use of historical data, but is tunable.	System does not make use of historical data, but is tunable.
The system shall be able to handle greater than 100,000 requests for matching per day.	Fulfills.	Fulfills.	Does not fulfill (limited to 64,000 records).	Fulfills.
The system shall support the needs of the medical groups as a primary concern.	Fulfills.	Fulfills.	Fulfills.	Fulfills.
The system shall provide facilities to estimate the training sample size needed for an initial tuning.	Does not fulfill; no training.	Does not fulfill; no training.	Does not fulfill; no training.	Does not fulfill; no training.
The system shall provide a means to review, override, and/or supplement its automated matching decisions.	Fulfills.	Fulfills.	Fulfills.	Fulfills.
The system shall provide reports on historical matching metrics.	Matching results may be saved for recall.	Matching results may be saved for recall.	Matching results may be saved for recall.	Matching results may be saved for recall.

Table 4. Features versus Patient-Matching Requirements (cont.)

Requirement	LinkageWiz	SureMatch	DeDupe4Excel	DataSet V
The system shall be able to supply a unique identifier for linked patients if requested.	Does not fulfill.	Fulfills.	Excel can generate.	Does not fulfill.
The system shall expose an application programming interface (API) for incorporation of matching functionality into general processing scripts.	Does not fulfill.	VB scripting available.	Does not fulfill directly.	Does not fulfill.
The system accuracy shall be able to reach 98% on average for a properly tuned system.	Undetermined.	Undetermined.	Undetermined.	Undetermined.
The system shall bear a total cost of ownership of less than $50K in the first year.	Fulfills.	Fulfills.	Fulfills.	Fulfills.

VII. Summary and Recommendations

The authors recommend trying several of these tools to determine which best meets a particular organization's needs and preferences.

ASSUMING EQUAL WEIGHTING OF THE RELEVANT features, the overall best products are DataSet V and LinkageWiz. For the price, these tools provide excellent ease of use, configurability of matching weights, ability to edit and export results, and user documentation. LinkageWiz is somewhat easier to learn and to use, whereas DataSet V provides somewhat greater configurability and a superior user interface for reviewing and editing match results.

SureMatch is also an effective tool for patient matching and provides the simplest user interface of all the tools. However, this simplicity comes at the cost of reduced functionality, because SureMatch provides no ability to modify the built-in weighting of match fields. Customization of these weightings can be very important in fine-tuning the matching performance based on characteristics of specific input data. Also, SureMatch can export data as a .csv file only, and provides limited user documentation. Finally, the demonstration version that was tested (v 6.2) included several bugs, although the vendor stated that a more stable version will be released in summer 2004.

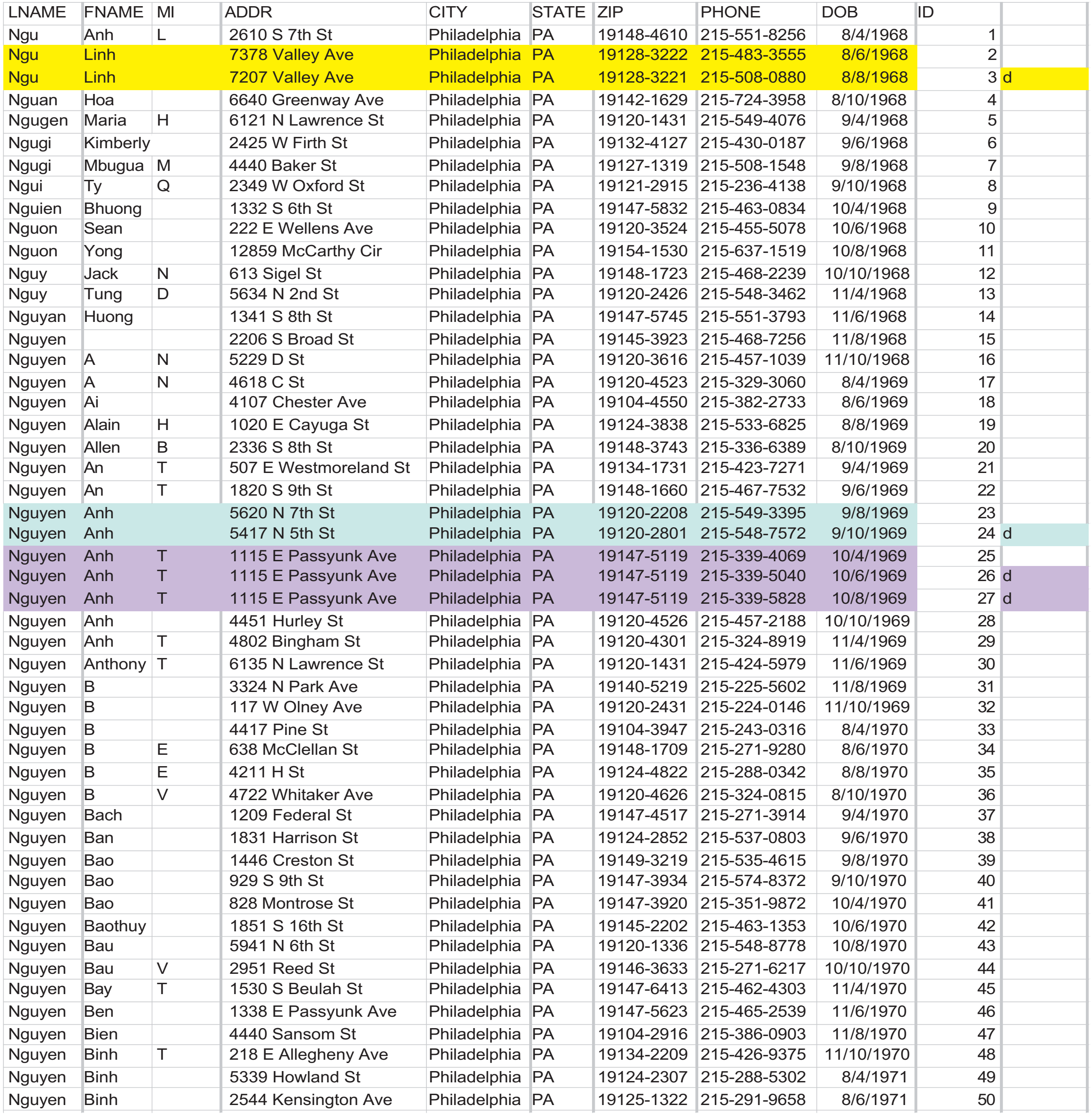
DeDupe4Excel, although effective in identifying matches and very low in price, has two significant limitations. First, the tool is designed solely to de-duplicate single data files, rather than match records from two distinct files. Although the user can concatenate two files to create a suitable input file for this tool, this requirement creates additional complexity and produces matching results that may be more difficult to process subsequently. More significantly, however, DeDupe4Excel is designed as an add-in to the Microsoft Excel spreadsheet application and, as such, has a data limit of about 65,000 records (i.e., the combined size of the files that the tool can process cannot exceed 65,000 records). Given the typical sizes of eligibility files and clinical data files, this limitation is prohibitive for all but the smallest provider organizations. Nevertheless, for small organizations with very limited budgets, this tool may be worth trying.

All of the tools reviewed apply probabilistic and fuzzy-matching techniques, and they have the ability to match records containing multiple identifying fields and possibly erroneous or omitted data values. Although use of the tools requires some familiarity with patient-matching concepts and some time spent learning and experimenting with the software, a competent data analyst should have no trouble learning to use the tools effectively. Given the variability in these tools' strengths and weaknesses, and the availability of demonstration copies, the authors recommend trying several of these tools to determine which best meets a particular organization's needs and preferences.

Appendix: Test Data Used for Product Evaluations

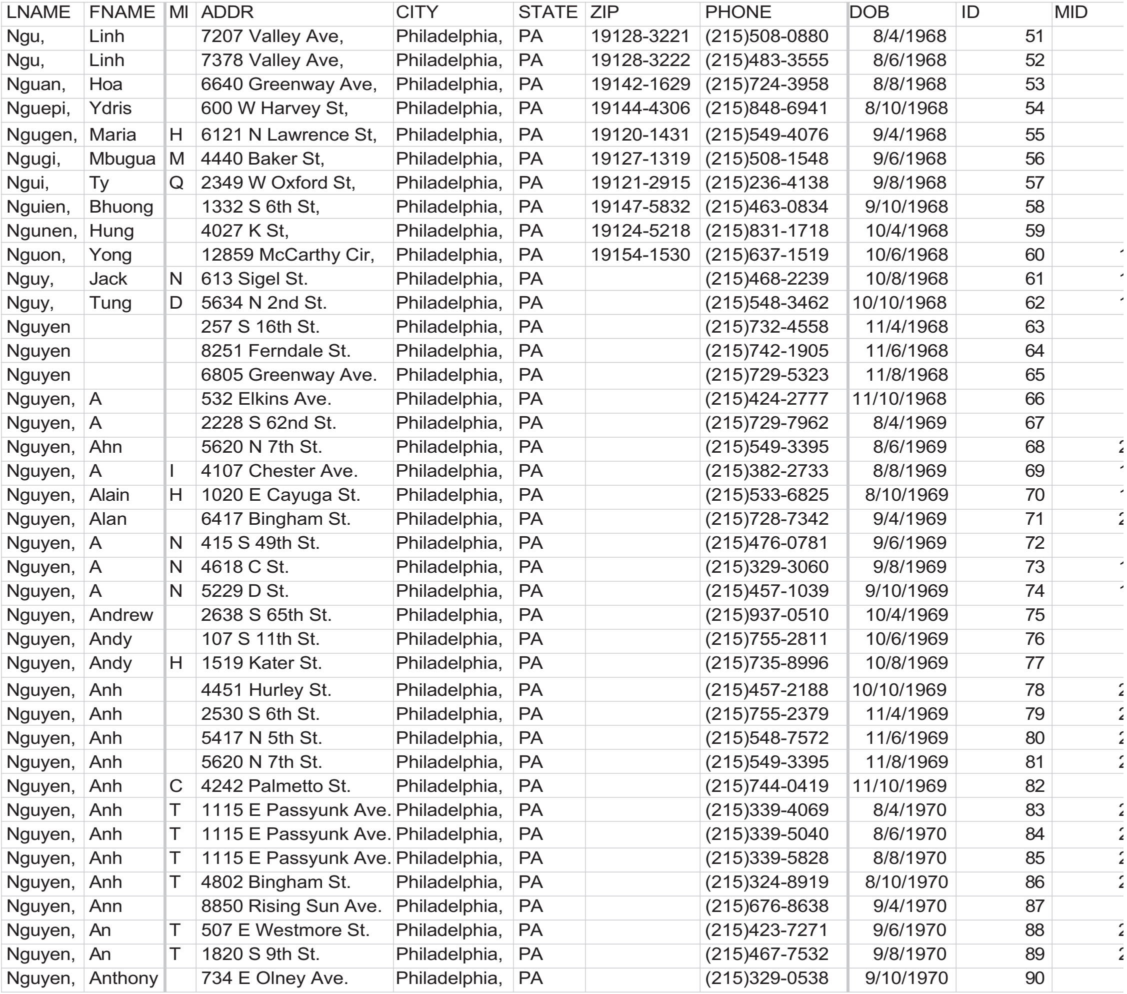
To assess each product's features through actual use, the researchers created a sample data set and processed these data with each tool. The data set consisted of a master file and a reference file, which were matched as part of the evaluation. Creating a small, synthetic data set provided a "gold standard" for assessing the effectiveness of each tool in detecting actual matches and ignoring non-matches. However, with this small, artificial data set, researchers were unable to: 1) extrapolate the findings with respect to matching accuracy to the data contents of the typical provider organization; or 2) assess the speed with which the tools can process files of the size typically matched by provider organizations.

Master file. Several duplicates were introduced into the master file to test the tools' ability to detect them. The groupings of duplicates in the test data are highlighted in color, with the records to be deleted flagged with a 'd.'



LNAME	FNAME	MI	ADDR	CITY	STATE	ZIP	PHONE	DOB	ID	
Ngu	Anh	L	2610 S 7th St	Philadelphia	PA	19148-4610	215-551-8256	8/4/1968	1	
Ngu	Linh		7378 Valley Ave	Philadelphia	PA	19128-3222	215-483-3555	8/6/1968	2	
Ngu	Linh		7207 Valley Ave	Philadelphia	PA	19128-3221	215-508-0880	8/8/1968	3	d
Nguan	Hoa		6640 Greenway Ave	Philadelphia	PA	19142-1629	215-724-3958	8/10/1968	4	
Ngugen	Maria	H	6121 N Lawrence St	Philadelphia	PA	19120-1431	215-549-4076	9/4/1968	5	
Ngugi	Kimberly		2425 W Firth St	Philadelphia	PA	19132-4127	215-430-0187	9/6/1968	6	
Ngugi	Mbugua	M	4440 Baker St	Philadelphia	PA	19127-1319	215-508-1548	9/8/1968	7	
Ngui	Ty	Q	2349 W Oxford St	Philadelphia	PA	19121-2915	215-236-4138	9/10/1968	8	
Nguien	Bhuong		1332 S 6th St	Philadelphia	PA	19147-5832	215-463-0834	10/4/1968	9	
Nguon	Sean		222 E Wellens Ave	Philadelphia	PA	19120-3524	215-455-5078	10/6/1968	10	
Nguon	Yong		12859 McCarthy Cir	Philadelphia	PA	19154-1530	215-637-1519	10/8/1968	11	
Nguy	Jack	N	613 Sigel St	Philadelphia	PA	19148-1723	215-468-2239	10/10/1968	12	
Nguy	Tung	D	5634 N 2nd St	Philadelphia	PA	19120-2426	215-548-3462	11/4/1968	13	
Nguyan	Huong		1341 S 8th St	Philadelphia	PA	19147-5745	215-551-3793	11/6/1968	14	
Nguyen			2206 S Broad St	Philadelphia	PA	19145-3923	215-468-7256	11/8/1968	15	
Nguyen	A	N	5229 D St	Philadelphia	PA	19120-3616	215-457-1039	11/10/1968	16	
Nguyen	A	N	4618 C St	Philadelphia	PA	19120-4523	215-329-3060	8/4/1969	17	
Nguyen	Ai		4107 Chester Ave	Philadelphia	PA	19104-4550	215-382-2733	8/6/1969	18	
Nguyen	Alain	H	1020 E Cayuga St	Philadelphia	PA	19124-3838	215-533-6825	8/8/1969	19	
Nguyen	Allen	B	2336 S 8th St	Philadelphia	PA	19148-3743	215-336-6389	8/10/1969	20	
Nguyen	An	T	507 E Westmoreland St	Philadelphia	PA	19134-1731	215-423-7271	9/4/1969	21	
Nguyen	An	T	1820 S 9th St	Philadelphia	PA	19148-1660	215-467-7532	9/6/1969	22	
Nguyen	Anh		5620 N 7th St	Philadelphia	PA	19120-2208	215-549-3395	9/8/1969	23	
Nguyen	Anh		5417 N 5th St	Philadelphia	PA	19120-2801	215-548-7572	9/10/1969	24	d
Nguyen	Anh	T	1115 E Passyunk Ave	Philadelphia	PA	19147-5119	215-339-4069	10/4/1969	25	
Nguyen	Anh	T	1115 E Passyunk Ave	Philadelphia	PA	19147-5119	215-339-5040	10/6/1969	26	d
Nguyen	Anh	T	1115 E Passyunk Ave	Philadelphia	PA	19147-5119	215-339-5828	10/8/1969	27	d
Nguyen	Anh		4451 Hurley St	Philadelphia	PA	19120-4526	215-457-2188	10/10/1969	28	
Nguyen	Anh	T	4802 Bingham St	Philadelphia	PA	19120-4301	215-324-8919	11/4/1969	29	
Nguyen	Anthony	T	6135 N Lawrence St	Philadelphia	PA	19120-1431	215-424-5979	11/6/1969	30	
Nguyen	B		3324 N Park Ave	Philadelphia	PA	19140-5219	215-225-5602	11/8/1969	31	
Nguyen	B		117 W Olney Ave	Philadelphia	PA	19120-2431	215-224-0146	11/10/1969	32	
Nguyen	B		4417 Pine St	Philadelphia	PA	19104-3947	215-243-0316	8/4/1970	33	
Nguyen	B	E	638 McClellan St	Philadelphia	PA	19148-1709	215-271-9280	8/6/1970	34	
Nguyen	B	E	4211 H St	Philadelphia	PA	19124-4822	215-288-0342	8/8/1970	35	
Nguyen	B	V	4722 Whitaker Ave	Philadelphia	PA	19120-4626	215-324-0815	8/10/1970	36	
Nguyen	Bach		1209 Federal St	Philadelphia	PA	19147-4517	215-271-3914	9/4/1970	37	
Nguyen	Ban		1831 Harrison St	Philadelphia	PA	19124-2852	215-537-0803	9/6/1970	38	
Nguyen	Bao		1446 Creston St	Philadelphia	PA	19149-3219	215-535-4615	9/8/1970	39	
Nguyen	Bao		929 S 9th St	Philadelphia	PA	19147-3934	215-574-8372	9/10/1970	40	
Nguyen	Bao		828 Montrose St	Philadelphia	PA	19147-3920	215-351-9872	10/4/1970	41	
Nguyen	Baothuy		1851 S 16th St	Philadelphia	PA	19145-2202	215-463-1353	10/6/1970	42	
Nguyen	Bau		5941 N 6th St	Philadelphia	PA	19120-1336	215-548-8778	10/8/1970	43	
Nguyen	Bau	V	2951 Reed St	Philadelphia	PA	19146-3633	215-271-6217	10/10/1970	44	
Nguyen	Bay	T	1530 S Beulah St	Philadelphia	PA	19147-6413	215-462-4303	11/4/1970	45	
Nguyen	Ben		1338 E Passyunk Ave	Philadelphia	PA	19147-5623	215-465-2539	11/6/1970	46	
Nguyen	Bien		4440 Sansom St	Philadelphia	PA	19104-2916	215-386-0903	11/8/1970	47	
Nguyen	Binh	T	218 E Allegheny Ave	Philadelphia	PA	19134-2209	215-426-9375	11/10/1970	48	
Nguyen	Binh		5339 Howland St	Philadelphia	PA	19124-2307	215-288-5302	8/4/1971	49	
Nguyen	Binh		2544 Kensington Ave	Philadelphia	PA	19125-1322	215-291-9658	8/6/1971	50	

Reference file. Matches are indicated with an entry in the MID (Master ID) column, indicating the ID of the record in the master file that this record matches. Records with no value in the MID field should not match any records in the master file. Note that the matches between the master and reference files were manually assigned, and the matching tools were tested to determine whether they could correctly detect these matches.



LNAME	FNAME	MI	ADDR	CITY	STATE	ZIP	PHONE	DOB	ID	MID
Ngu,	Linh		7207 Valley Ave,	Philadelphia,	PA	19128-3221	(215)508-0880	8/4/1968	51	
Ngu,	Linh		7378 Valley Ave,	Philadelphia,	PA	19128-3222	(215)483-3555	8/6/1968	52	
Nguan,	Hoa		6640 Greenway Ave,	Philadelphia,	PA	19142-1629	(215)724-3958	8/8/1968	53	
Nguepi,	Ydris		600 W Harvey St,	Philadelphia,	PA	19144-4306	(215)848-6941	8/10/1968	54	
Ngugen,	Maria	H	6121 N Lawrence St,	Philadelphia,	PA	19120-1431	(215)549-4076	9/4/1968	55	
Ngugi,	Mbugua	M	4440 Baker St,	Philadelphia,	PA	19127-1319	(215)508-1548	9/6/1968	56	
Ngui,	Ty	Q	2349 W Oxford St,	Philadelphia,	PA	19121-2915	(215)236-4138	9/8/1968	57	
Nguien,	Bhuong		1332 S 6th St,	Philadelphia,	PA	19147-5832	(215)463-0834	9/10/1968	58	
Ngunen,	Hung		4027 K St,	Philadelphia,	PA	19124-5218	(215)831-1718	10/4/1968	59	
Nguon,	Yong		12859 McCarthy Cir,	Philadelphia,	PA	19154-1530	(215)637-1519	10/6/1968	60	1
Nguy,	Jack	N	613 Sigel St.	Philadelphia,	PA		(215)468-2239	10/8/1968	61	1
Nguy,	Tung	D	5634 N 2nd St.	Philadelphia,	PA		(215)548-3462	10/10/1968	62	1
Nguyen			257 S 16th St.	Philadelphia,	PA		(215)732-4558	11/4/1968	63	
Nguyen			8251 Ferndale St.	Philadelphia,	PA		(215)742-1905	11/6/1968	64	
Nguyen			6805 Greenway Ave.	Philadelphia,	PA		(215)729-5323	11/8/1968	65	
Nguyen,	A		532 Elkins Ave.	Philadelphia,	PA		(215)424-2777	11/10/1968	66	
Nguyen,	A		2228 S 62nd St.	Philadelphia,	PA		(215)729-7962	8/4/1969	67	
Nguyen,	Ahn		5620 N 7th St.	Philadelphia,	PA		(215)549-3395	8/6/1969	68	2
Nguyen,	A	I	4107 Chester Ave.	Philadelphia,	PA		(215)382-2733	8/8/1969	69	1
Nguyen,	Alain	H	1020 E Cayuga St.	Philadelphia,	PA		(215)533-6825	8/10/1969	70	1
Nguyen,	Alan		6417 Bingham St.	Philadelphia,	PA		(215)728-7342	9/4/1969	71	2
Nguyen,	A	N	415 S 49th St.	Philadelphia,	PA		(215)476-0781	9/6/1969	72	
Nguyen,	A	N	4618 C St.	Philadelphia,	PA		(215)329-3060	9/8/1969	73	1
Nguyen,	A	N	5229 D St.	Philadelphia,	PA		(215)457-1039	9/10/1969	74	1
Nguyen,	Andrew		2638 S 65th St.	Philadelphia,	PA		(215)937-0510	10/4/1969	75	
Nguyen,	Andy		107 S 11th St.	Philadelphia,	PA		(215)755-2811	10/6/1969	76	
Nguyen,	Andy	H	1519 Kater St.	Philadelphia,	PA		(215)735-8996	10/8/1969	77	
Nguyen,	Anh		4451 Hurley St.	Philadelphia,	PA		(215)457-2188	10/10/1969	78	2
Nguyen,	Anh		2530 S 6th St.	Philadelphia,	PA		(215)755-2379	11/4/1969	79	2
Nguyen,	Anh		5417 N 5th St.	Philadelphia,	PA		(215)548-7572	11/6/1969	80	2
Nguyen,	Anh		5620 N 7th St.	Philadelphia,	PA		(215)549-3395	11/8/1969	81	2
Nguyen,	Anh	C	4242 Palmetto St.	Philadelphia,	PA		(215)744-0419	11/10/1969	82	
Nguyen,	Anh	T	1115 E Passyunk Ave.	Philadelphia,	PA		(215)339-4069	8/4/1970	83	2
Nguyen,	Anh	T	1115 E Passyunk Ave.	Philadelphia,	PA		(215)339-5040	8/6/1970	84	2
Nguyen,	Anh	T	1115 E Passyunk Ave.	Philadelphia,	PA		(215)339-5828	8/8/1970	85	2
Nguyen,	Anh	T	4802 Bingham St.	Philadelphia,	PA		(215)324-8919	8/10/1970	86	2
Nguyen,	Ann		8850 Rising Sun Ave.	Philadelphia,	PA		(215)676-8638	9/4/1970	87	
Nguyen,	An	T	507 E Westmore St.	Philadelphia,	PA		(215)423-7271	9/6/1970	88	2
Nguyen,	An	T	1820 S 9th St.	Philadelphia,	PA		(215)467-7532	9/8/1970	89	2
Nguyen,	Anthony		734 E Olney Ave.	Philadelphia,	PA		(215)329-0538	9/10/1970	90	