

# **National Practice Assessment: Immunization Information System Patient De-duplication**

**Frederic Grant, PhD, MPH**  
**September 2012**

# Agenda

- ❑ **Background**
  - Objectives
  - Focus Areas
- ❑ **Methods**
- ❑ **Literature Review**
- ❑ **Results**
- ❑ **Discussion**
- ❑ **Questions**

# Project Background

- ❑ **CDC Patient-level De-duplication Panel formed 2011**
- ❑ **External sources for patient records are increasing**
  - HITECH Act
  - Meaningful Use
  - Data Interoperability / Health Information Exchanges
- ❑ **Duplicate patient records undermine credibility**
  - Data quality
  - Accuracy of immunization status
  - Patient care

# Five Areas of Focus

Five Areas	Goals
New robust IIS patient-level de-duplication test cases	<ul style="list-style-type: none"> <li>• Tools to improve patient-level de-duplication practices</li> <li>• Update of 2002 CDC test cases to best practice standards</li> </ul>
Practice-based solutions for evaluating IIS patient matching and de-duplication approaches	<ul style="list-style-type: none"> <li>• National Practice Assessment</li> <li>• Validation of contextual models</li> <li>• Identification of best practices</li> <li>• Definition of common vocabulary</li> <li>• Determination of emerging role of the Master Patient Index (MPI)</li> <li>• Identification of sensitivity and specificity and other measures</li> </ul>
Methods supporting manual data entry and incoming data	<ul style="list-style-type: none"> <li>• Problems and solutions</li> <li>• Guidance on pre-screening incoming records</li> </ul>
Methods supporting the examination of existing data	<ul style="list-style-type: none"> <li>• Problems and solutions</li> <li>• Guidance on retrospective processing</li> </ul>
Manual review practices	<ul style="list-style-type: none"> <li>• Problems and solutions</li> <li>• Merge and un-merge capabilities</li> </ul>

# National Practice Assessment Focus Areas

Area	Assessment Objective
National practice characterization	<ul style="list-style-type: none"> <li>• Overall capabilities, needs, and degree of automation</li> </ul>
De-duplication software	<ul style="list-style-type: none"> <li>• Origins of de-duplication software</li> <li>• Degree of satisfaction</li> <li>• Plans for replacement</li> <li>• Architecture – where de-duplication logic resides</li> <li>• Algorithmic capabilities</li> <li>• Status of MPI integration</li> </ul>
Causes of patient duplicate records	<ul style="list-style-type: none"> <li>• Problems and solutions</li> <li>• Test case generation</li> <li>• Guidance on pre-screening incoming records</li> </ul>
Methods supporting de-duplication of manual, incoming, and existing data	<ul style="list-style-type: none"> <li>• Trends, patterns, needs</li> </ul>
Ability to detect specific types of errors	<ul style="list-style-type: none"> <li>• Twins, typos, misspellings, transpositions</li> <li>• Data field usage</li> </ul>
Data usage for de-duplication purposes	

# Publication Manuscript

## Abstract

The purpose of this National Practice Assessment was to gather information about the patient-level immunization information system (IIS) de-duplication software, procedures, tools, problems, and practices.

The information collected from this survey is being used by the CDC De-duplication Expert Panel.

## Focus

Patient matching and patient de-duplication

## Goals

Better understanding of the needs of the IIS national practice community and update 2002 test cases

## Results

Insights into software, procedures, tools, practices, problems, and trends

# Literature Review Highlights

Arzt, N. H. (2008). *Architecture for Person Matching and De-duplication*.

Grannis, S. J., Overhage, J. M., & McDonald, C. J. (2003). *Analysis of a probabilistic record linkage technique without human review*

Public Health Informatics Institute (PHII) , (2006) The Unique Records Portfolio. Decatur, GA: Public Health Informatics Institute. Clyde & Salkowitz, 2006.

HIMSS. (2009). Patient Identity Integrity.

Williams W, Lowery NE, Lyalin D, Lambrecht N, Riddick S, Sutcliff C, Papadouka V. (2011). Development and utilization of best practice operational guidelines for immunization information systems.

## □ Paucity of peer reviewed research on patient de-duplication specific to IIS

## □ Described by various terminologies

- patient identity management
- patient matching
- master data management
- duplicate detection
- record linking
- identity resolution
- fuzzy duplication detection
- entity matching
- patient identity integrity

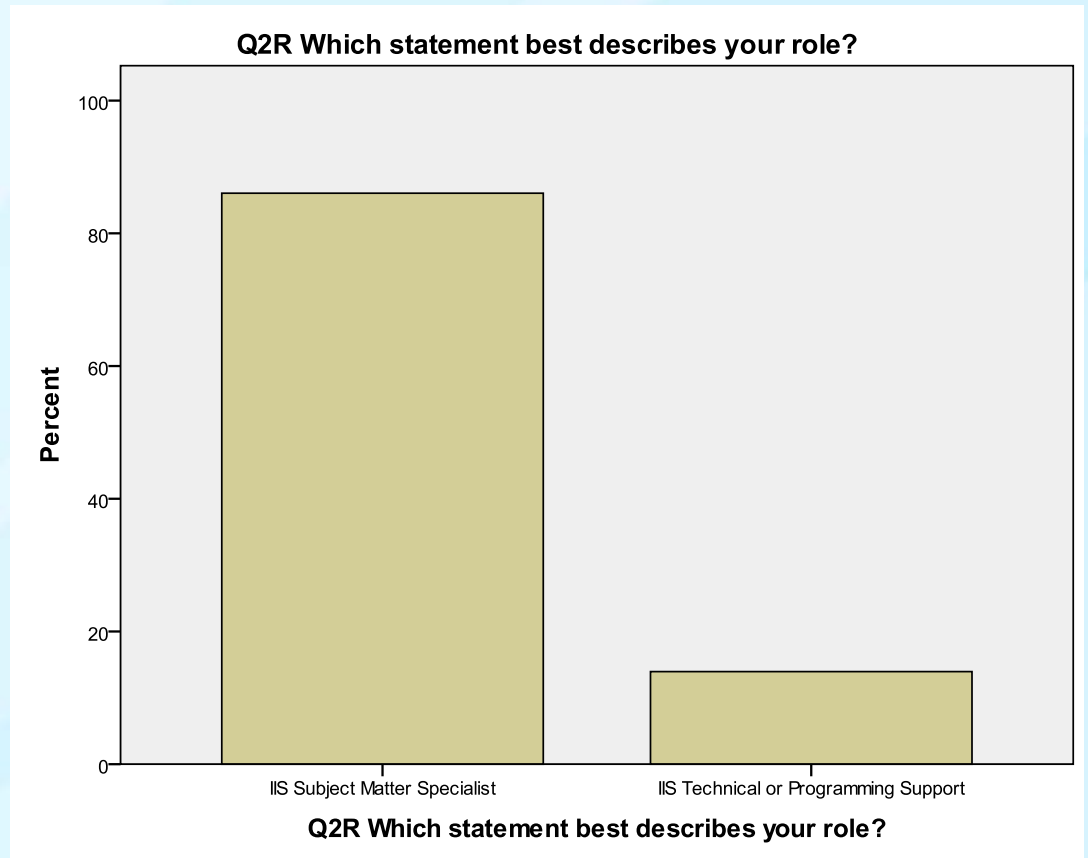
## □ IIS represents an important focus of Meaningful Use

# Methods

- ❑ **Web-based SurveyMonkey survey with 22 questions**
- ❑ **Developed and refined by CDC Expert Panel**
- ❑ **Delphi approach – structured facilitation**
- ❑ **Piloted with CDC Expert Reviewers**
- ❑ **Quantitative and qualitative inputs**
- ❑ **Structured and unstructured data responses**

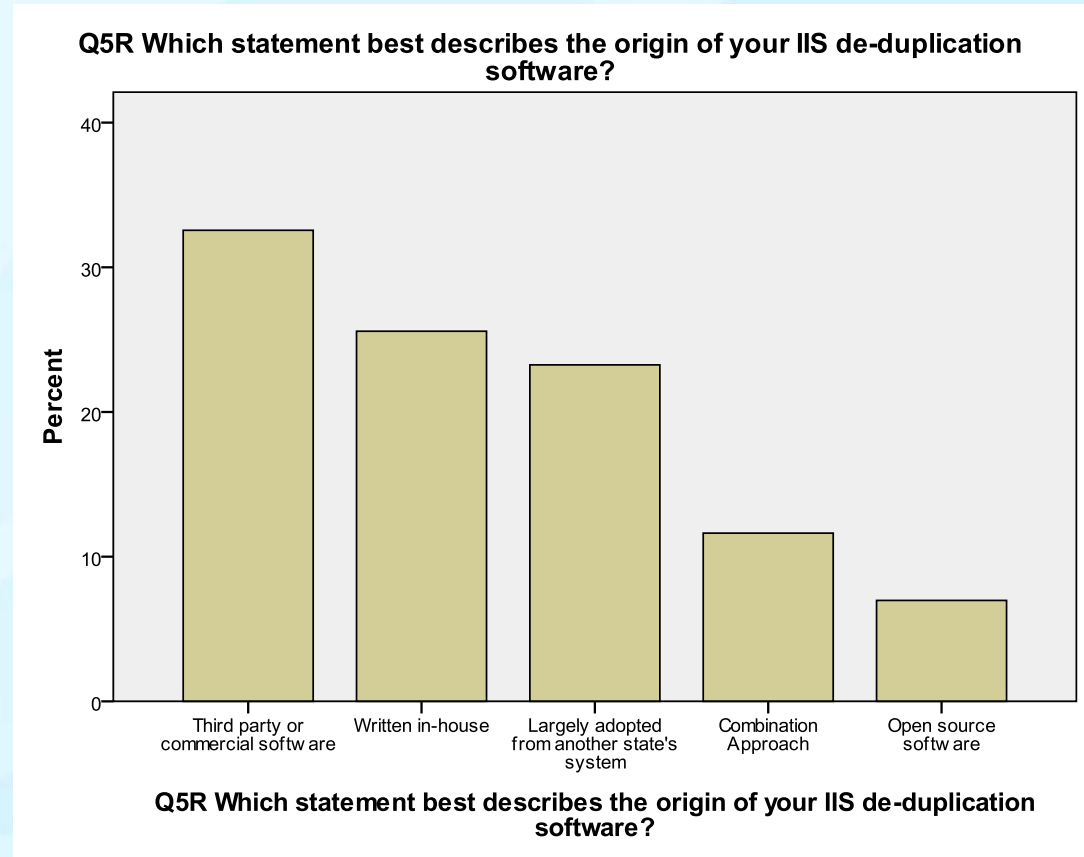
# Respondent Information

- Target was State and Territorial IIS implementations
- 43 respondents
  - 86% SME
  - 14% Technical
- Manage an average of 4.7 million patient records
- On average, over 345,000 patient records added annually



# Origins of Software

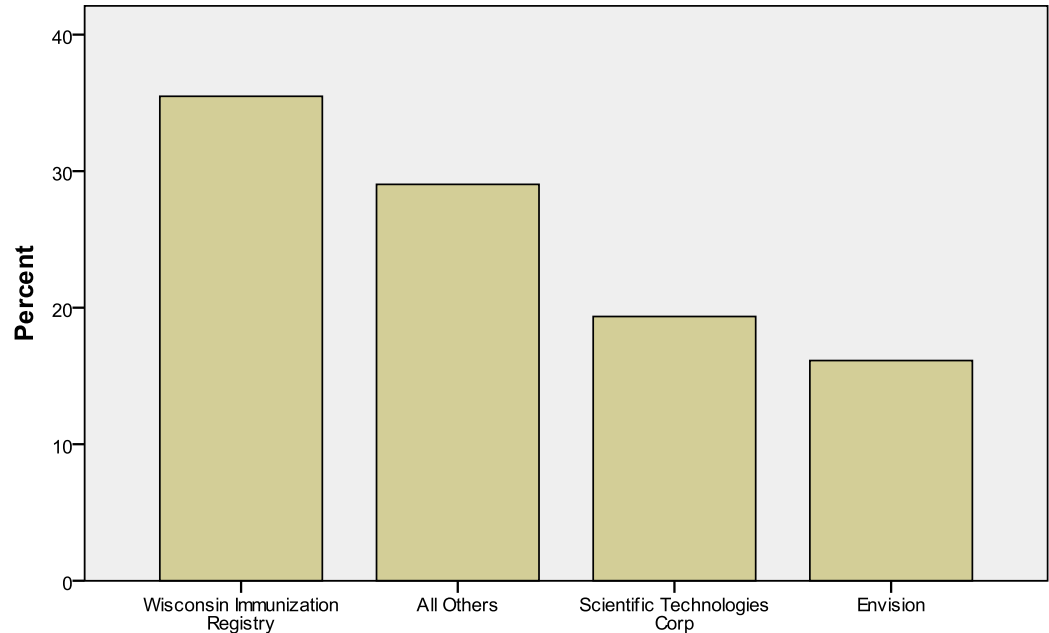
- Third party or commercial - 14
- Written in-house - 11
- Largely adopted from another state's system - 10
- Combination approach - 5
- Open source software - 3
- Total - 43



# Software Source

- Wisconsin Immunization Registry -11
- Scientific Technologies Corp - 6
- Envision - 5
- All Others - 9
- Total - 31
- Missing -12
- Total - 43

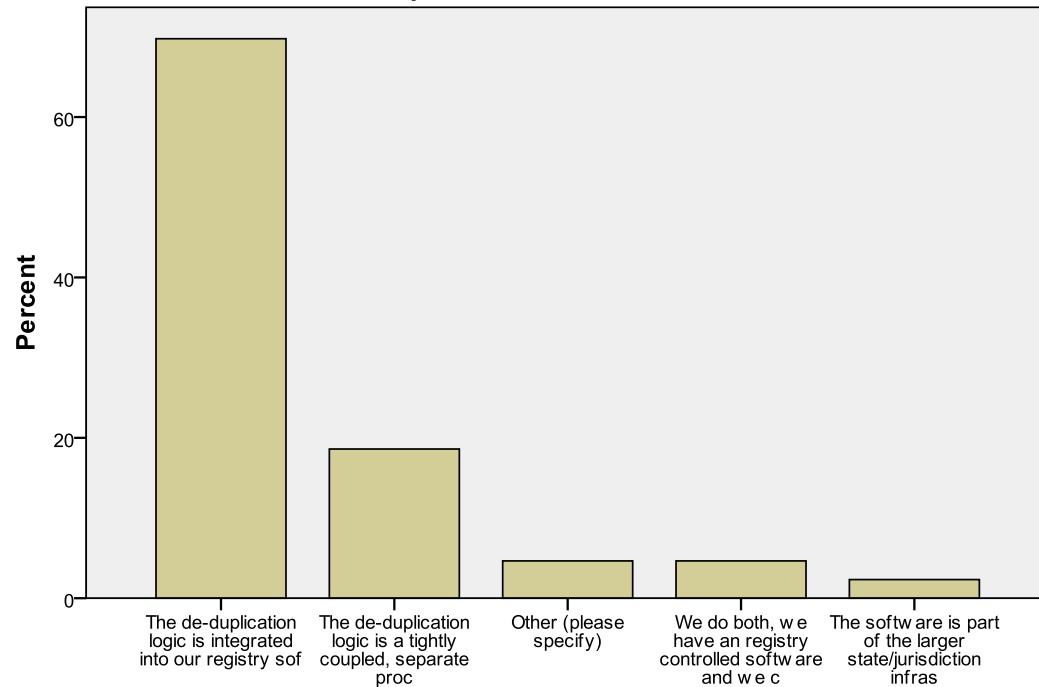
**Q6R If your IIS de-duplication process is based on third party, commercial software, open source, or another state's software, please indicate the name of the software, vendor, or origin source of your software.**



**Q6R If your IIS de-duplication process is based on third party, commercial software, open source, or another state's software, please indicate the name of the software, vendor, or origin source of your software.**

# Where De-duplication Logic Resides

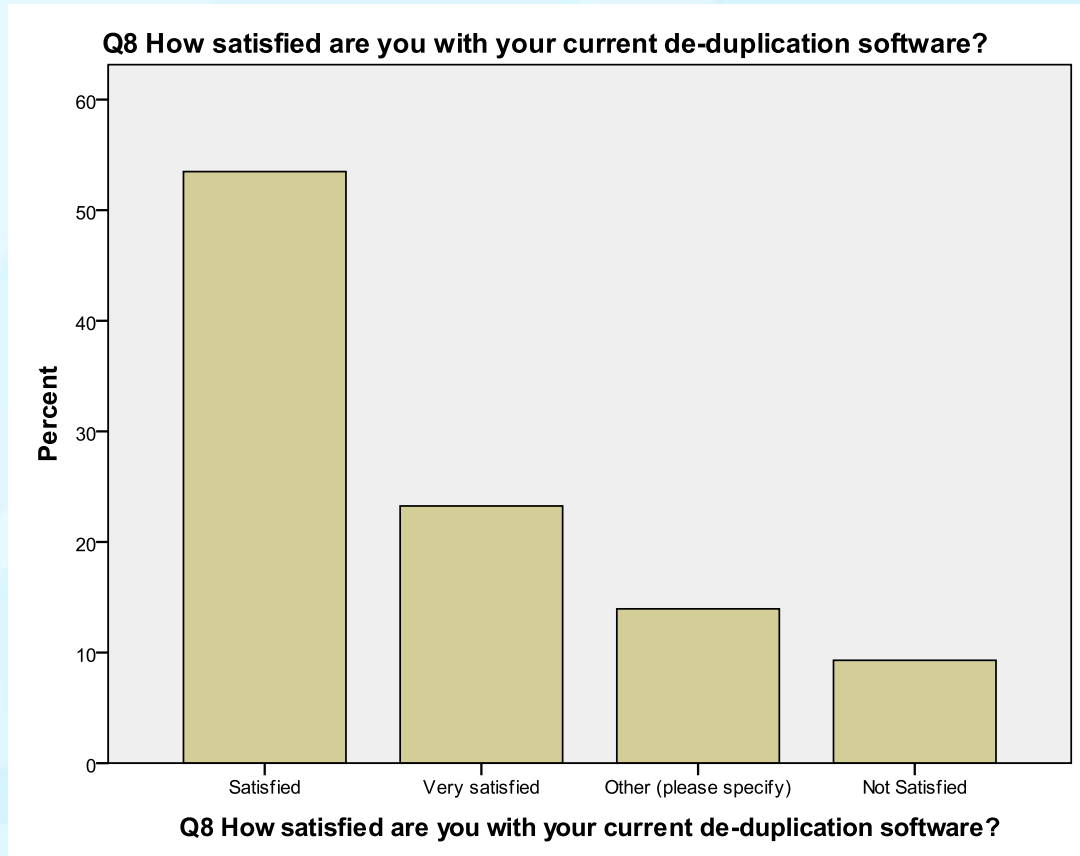
**Q7 Which statement best describes where the software for your de-duplication process resides?**



**Q7 Which statement best describes where the software for your de-duplication process resides?**

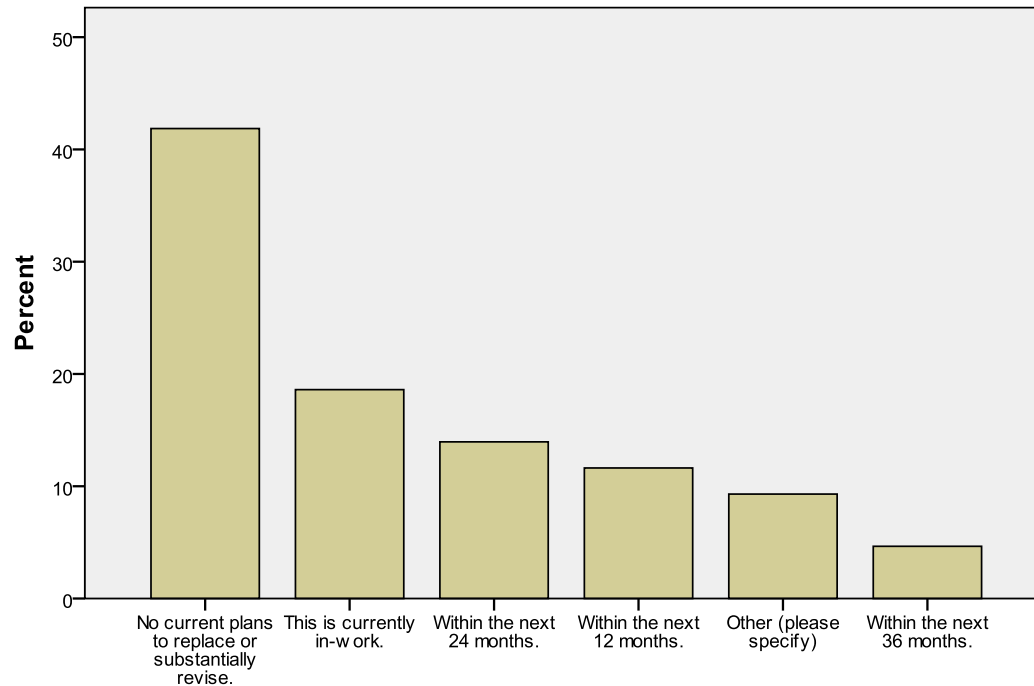
# Satisfaction with Software

- Satisfied - 53.5%
- Very satisfied - 23.3%
- Not satisfied - 9.3%
- Replacement of de-duplication software currently in-process - 18.6%
- Plans to replace or substantially revise software within the next 36 months or sooner - 30%



# De-duplication Replacement Plans

**Q9 When do you plan to replace or substantially revise your de-duplication software?**



**Q9 When do you plan to replace or substantially revise your de-duplication software?**

# Discussion

## ❑ Key themes

- Life changes reveal system weaknesses
- Optional data such as Social Security Number or Medical Record Number is of high value
- Greater efficiency and consistency in manual review is essential; experience and training make a difference
- Systems need to provide functionality around efficient merging and un-merging of patient records
- Circumstances and resources vary considerably

## ❑ Potential significant impact through establishment of national standards

## Discussion Continued

### ❑ Best results with hybrid algorithmic approaches

- 53.5% - Detect phonetic variations in names
  - 81.4% - Detect duplicates in hyphenated name
  - 81.4% - Employ a name-matching algorithm that recognizes nicknames as matches (e.g., Robert = Bob).
- 74.4% - Edit distance and phonetic name-matching algorithms (e.g., Soundex, NYSIIS, Metaphone, etc.) and/or similar types of algorithms used
- Some respondents had ability to determine probabilistic matches on records but lacked ability to detect character typos and transpositions
- 48.8% - Standardize patient addresses for matching purposes
- 95.3% - Require a complete date of birth for patient matching purposes
- 55.8% - Detect typographical errors in birthdates
- 72.1% - Take precautions to prevent the false matching of twins

## Discussion Continued

- ❑ **97.7% - Processes in place that require staff to regularly review records that have been flagged as possible duplicates**
  - 51.2% - Formal published procedures for evaluation of potential duplicate records
  - 83.7% - Review process takes more time than they would like
  
- ❑ **Tolerance for duplicate patient records**
  - Mean 3.81% (s.d. = 4.268)
  - Median 2.5%.
  - Consistent with expert panel experience

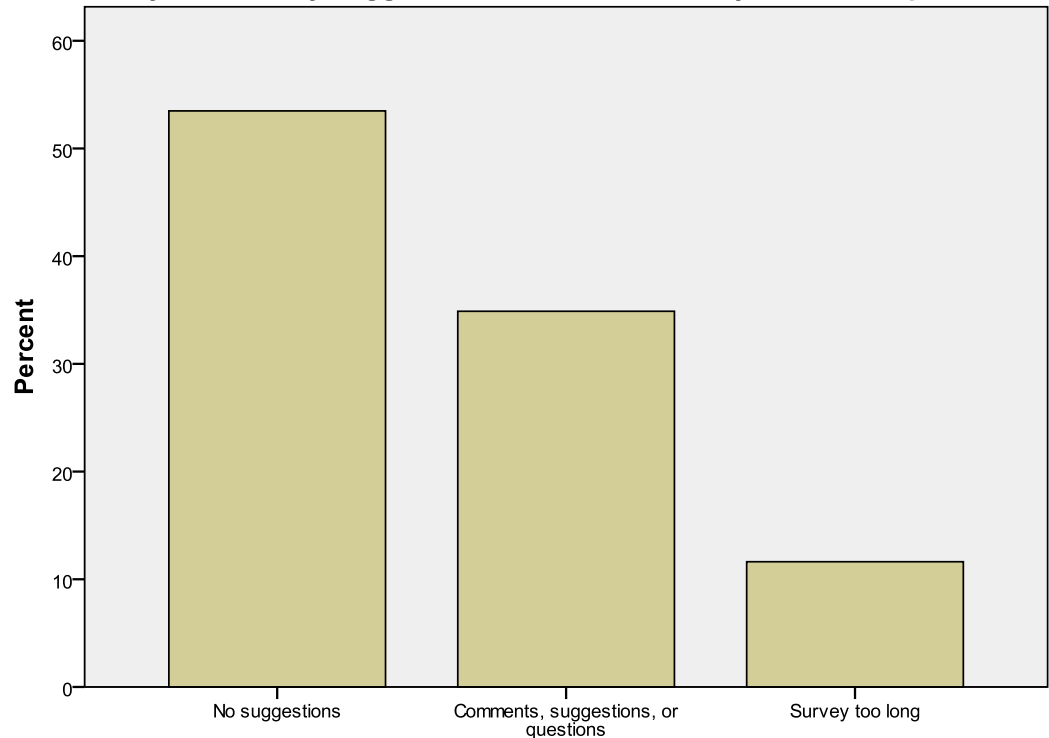
# Conclusion

- ❑ Important but understudied area
- ❑ Significant expense and complexity
- ❑ Sustained investment required for continuous improvement
- ❑ Close monitoring of ONC and other governmental policy decisions regarding levels of accuracy required
- ❑ Potential significant impact through establishment of national standards

# Improving Future Surveys

- Reduce number of questions
- Include survey in final report
- Include types of questions in annual survey

Q22R Do you have any suggestions on how this survey could be improved?



Q22R Do you have any suggestions on how this survey could be improved?

## Next Steps

- ❑ **Patient de-duplication test case generation – 8 categories**
- ❑ **Final report production and distribution**
- ❑ **Continued learning**

# Study Limitations

- ❑ Web survey-based studies have limitations regarding study design
- ❑ Few previous surveys have specifically examined IIS patient-level de-duplication practices
- ❑ Development of the survey instrument was relative to the informational needs of the CDC Expert Panel
- ❑ Public Health agencies were the key participants in this investigation
- ❑ Some respondents reported that the survey was lengthy
- ❑ It is possible that some jurisdictions were unable to respond due to workloads
- ❑ Further study of national IIS patient de-duplication challenges, issues, and practices is needed

## Acknowledgements

- ❑ Special thanks to Warren Williams, CDC; Stuart Myerburg, CDC; David Lyalin, CDC; Jennifer Wain, NG; Eric Larson, NG; Celia Toles, NG; Lindsay Ryan, NG
- ❑ Expert Panel Coauthor Information: Michael Berry / Rhode Island; Gerry Bragg / Michigan; Nathan Bunker / Utah; Shaun Grannis, MD / Indiana; Rick Hall, PhD / NY; Steve Jarvis / Colorado; Brian Jorgage / Philadelphia; Mary Beth Kurilo, MPH/MSW / Oregon; Linda Leubchow / Minnesota; Christie Levy / Mississippi; Megan Meldrum / NY; Chris Pratt / Utah; Helen Redfield / Texas; Cecile Town / New Mexico.

# Questions

**Jennifer Wain**  
**[jua7@cdc.gov](mailto:jua7@cdc.gov)**  
**678-530-8841**

**Stuart Myerburg**  
**[jyz0@cdc.gov](mailto:jyz0@cdc.gov)**  
**404-639-1813**

**Frederic Grant**  
**[hlc4@cdc.gov](mailto:hlc4@cdc.gov)**  
**770-262-7593**

**For more information please contact Centers for Disease Control and Prevention**

1600 Clifton Road NE, Atlanta, GA 30333

Telephone, 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348

E-mail: [cdcinfo@cdc.gov](mailto:cdcinfo@cdc.gov) Web: [www.cdc.gov](http://www.cdc.gov)