



## Link Plus

# *Probabilistic Record Linkage Software from CDC*

**SAFER • HEALTHIER • PEOPLE™**

AL • CHRONIC DISEASE PREVENTION • INFECTIOUS DISEASE PROTECTION • IMMUNIZATION • INJURY PREVENTION • PRI  
CHILD HEALTH • GLOBAL PARTNERSHIPS • MINORITY OUTREACH • MONITORING HEALTH • COMMUNITY PARTNERSHIPS  
SAFER • HEALTHIER • PEOPLE • SAFER • HEALTHIER • PEOPLE • SAFER • HEALTHIER • PEOPLE • SAFER • HEA  
HEALTH EDUCATION • EPIDEMIOLOGY • WORKPLACE HEALTH • IMMUNIZATION • WORKPLACE SAFETY • TRAINING • C  
LTH • PREVENTION RESEARCH • PRIVATE SECTOR PARTNERSHIPS • PUBLIC HEALTH WORKFORCE • WOMEN'S HEALTH

# Today's Presentation

- Context
- ***Summary description of the program***
- ***Brief overview of record-linkage principles***
- ***Tour of the program's features***
- ***Demonstration***

# Cancer Registries

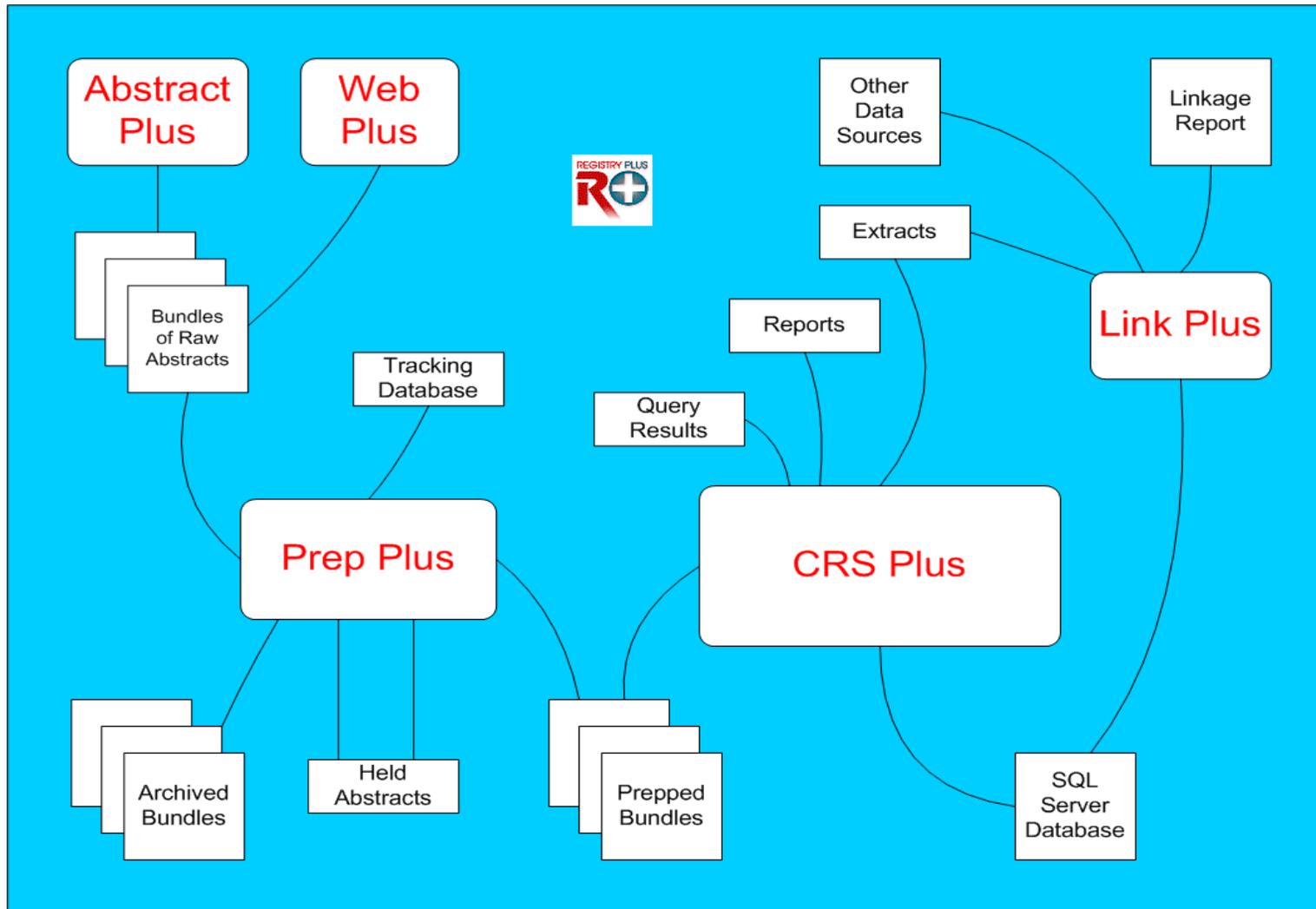
- All states have central cancer registries
- Most of them participate in the National Program of Cancer Registries
- State laws require diagnostic and treatment facilities to report most kinds of cancer to their central registry.

# NPCR



- Established by US Congress in 1992
  - Funding
  - Training
  - Standard requirements

# Registry Plus



# Consensus Standards

- North American Association of Central Cancer Registries
  - State registries
  - National institutions
  - Other interested parties



# NAACCR Records

- Nearly 400 data items
  - Demographic
  - Cancer Identification
  - Diagnosis
  - Treatment
  - Supporting Text

# Link Plus

- Useful
- Free software
- Easy to use
- Interesting

# Link Plus Is Useful

- Essential Cancer Registry Tasks:
  - De-duplication
  - Death Clearance
  - Case Finding
  - Special Studies

# Link Plus Is Free

- \$0.00

# Link Plus Is Easy To Use

- Designed especially for cancer registry work
- Mathematics largely hidden from user
- Practical default values supplied for many tasks
- Familiar Windows interface
- Includes help and samples

# Link Plus Is Interesting

- Program written by a mathematical statistician
- Specifications based on research into the published literature
- Tested by researchers experienced in record-linkage
- Results are clear and accessible to non-specialists

# Record-Linkage Concepts

$$P(\gamma | M) = \prod_{i=1}^K m_i^{\gamma^i} (1-m_i)^{(1-\gamma^i)}$$

and

$$P(\gamma | U) = \prod_{i=1}^K u_i^{\gamma^i} (1-u_i)^{(1-\gamma^i)} .$$

# Record-Linkage Concepts

- Find the records in File A that seem to match records in File B
- Calculate a score that indicates, for any pair of records, how likely it is that they both refer to the same person
- Discard unlikely matched pairs (low scores)
- Sort the likely and possible matched pairs in order of their scores
- Visually review a range of uncertain matches

# Record-Linkage Concepts

- Blocking
  - Large files can make impossible resource demands
  - Discard very unlikely record-pairings from the start

# Record-Linkage Concepts

- The total score of a linkage for any two records is the sum of the scores from matching individual fields
- The score assigned to a matching of individual fields is based on
  - The probability that the fields will agree between records that truly refer to the same person
  - Reduced by the probability that they will by chance agree between records that are not a true match.

# Record-Linkage Concepts

- Comparators
  - Find partial, approximate, or fuzzy matches
  - Value of match on a particular field can be other than “yes” or “no”, 1 or 0.

# Record-Linkage Concepts

- Probabilistic weights
  - Field-specific – birthdate versus sex
  - Value-specific - “William” versus “Artemis”

# Record-Linkage Concepts

- De-duplication is a special case of record linkage.
- Records in the same file are blocked, compared, and scored against each other.
- The result is a ranked list of record pairs.
- High-scoring pairs may be duplicates.

# Demonstration

- Synthetic data
- 50000 records of simulated cancer registry data
- 10000 records of bogus death certificate data, including:
  - 100 records that should match records in the cancer data, of which:
    - 20 have multiple errors

# Link Plus Today

- We have looked at a version of Link Plus that is scheduled for release in May.
- The version currently available does not contain the “Clerical Review” features.
- The program will soon be subjected to usability and interface design review



# Obtaining Link Plus



Link Plus can be downloaded from

<http://www.cdc.gov/cancer/npcr/>

**SAFER • HEALTHIER • PEOPLE™**

AL • CHRONIC DISEASE PREVENTION • INFECTIOUS DISEASE PROTECTION • IMMUNIZATION • INJURY PREVENTION • PRI  
CHILD HEALTH • GLOBAL PARTNERSHIPS • MINORITY OUTREACH • MONITORING HEALTH • COMMUNITY PARTNERSHIPS  
SAFER • HEALTHIER • PEOPLE • SAFER • HEALTHIER • PEOPLE • SAFER • HEALTHIER • PEOPLE • SAFER • HEA  
HEALTH EDUCATION • EPIDEMIOLOGY • WORKPLACE HEALTH • IMMUNIZATION • WORKPLACE SAFETY • TRAINING • C  
LTH • PREVENTION RESEARCH • PRIVATE SECTOR PARTNERSHIPS • PUBLIC HEALTH WORKFORCE • WOMEN'S HEALTH