Record Linkage

- **Refers to:**
  - Task of finding records in a data set that refer to the same entity across different data sources.
    - Example: Admissions & Path report @ hospital
    - Example: Cancer Abstract & Death certificate @ state registry
  - Joining datasets based on entities that may or may not share a common identifier.
    - Example: @ State Registry: SSN

- **Definition:**
  - Joining records from one data source with another that describes the same entity.
  - Narrow focus of identifying matching entities across different datasets
    - Patient
    - Tumor
Record Linkage

- Deterministic
  - All or some of identifiers are identical
    - SSN, Name, DOB, etc.

- Probabilistic
  - Fuzzy matching
  - Weights are computed based on identifier's ability to correctly identify match – weights are used to calculate probability that records represent same entity.
    - Above certain probability considered match
    - Below certain probability considered non-match
    - In between reviewed manually

FCDS Record Linkage

- Variables used
  - SSN
  - First Name
  - Last Name
  - Date of Birth
  - Sex
  - County of residence

- Race
- Ethnicity
- Address
- Zip Code
- Telephone
- Maiden Name
- Middle Initial
- Alias
Record Linkages
• First outside linkage done in 1995

Number of linkages
o 2000-2005: 1-5 per year
o 2005-2010: 5-10 per year
o 2010+: 10+ per year

Time intensive
o Especially manual review - possibilities

Notable Linkages
• World Trade Center Health Registry Linkage
  o 3 linkages so far since 2008
• Cancer Risk among Firefighters and Emergency Service Rescuers and Officers Exposed to the World Trade Center Disaster
  o 2 linkages to date since 2010
• Cancer in WTC Responders participating in the WTC Medical Monitoring and Treatment Program
  o 1 linkage to date since 2011
• Camp Lejeune Health Survey
  o Assess whether there is an association between exposure to contaminated water at Camp Lejeune and cancer and other specified health conditions
  o 1 linkage to date since 2013

Notable Linkages
• Cancer Risk in Solid Organ Transplant Recipients: The Transplant Cancer Match Study
  o 1 linkage to date 2011
• Southern Community Cohort Study
  o 100,000 adults residents across the South for long term follow-up to evaluate determinants of cancer and other chronic diseases
  o 4 linkages to date since 2009
• Nurses Health Study
  o Largest & longest running investigation of factors that influence women’s health: 228,000 participants
  o 26 linkages to date since 2005
• Health Professionals Follow-up Study
  o All male study to evaluate nutritional factors and relationship to the incidence of serious illnesses—cancer being one of the illnesses
  o 10 linkages since 2003
Notable Linkages

- **4 Different linkages with NIOSH 2011 & 2012**
  - Cancer Incidence and mortality among workers exposed to Polychlorinated Biphenyls (PCBs) in Electrical Capacitor Manufacturing Plants
  - Bladder Cancer among workers exposed to o-Toluidine in chemical manufacturing plant
  - Cancer risk among U.S. Firefighters
  - Flight attendants Health Study: Incidence of breast cancer and other cancers among Female Flight Attendants

- **Forteo Patient Registry**
  - Estimate incidence of osteosarcoma in patients who have received treatment with Forteo – a biosynthetic human parathyroid hormone for women with osteoporosis and men with low bone mineral density.
  - 4 linkages to date since 2011

- **Cancer Epidemiology in Adventists**
  - Cohort of 96,000 participants to address broad question of dietary factors that reduce or increase the risk of common cancers.
  - 2 linkages to date since 2011

- **Infertility Follow-up study**
  - Follow-up on cohort of 12,000+ women to assess cancer risk in relation to causes of infertility and therapeutic regimens used to treat these causes
  - 1 linkage to date since 2011

- **Black Women’s Health Study**
  - Evaluate causes and preventives of cancers and other serious illnesses in African-American women
  - 5 Linkages to date since 2005

- **HIV/AIDS registry match**
  - 2 linkages since 2001

Notable Linkages (in progress)

- **Assisted Reproductive Technology & Risk of Childhood Cancer**
  - Compare incidence of childhood cancer in children with assisted conception to that in the general population

- **Prostate, Lung, Colorectal and Ovarian (PLCO) Screening Trial**
  - Whether screening with flexible sigmoidoscopy can ↓ mortality from colorectal cancer, whether screening with chest X-ray can ↓ mortality from lung cancer, whether screening men with digital rectal examination (DRE) plus serum prostate-specific antigen (PSA) can ↓ mortality from prostate cancer, and whether screening women with CA125 and transvaginal ultrasound (TVU) can ↓ mortality from ovarian cancer
Conclusion

- Good demographics vital
  - reliability of match
  - time involved

- Affects quality of match
  - Outside
  - Internal

- Linkages are vital in studying cancer etiology

$