CDC Paper Trail & Simulation

Overview

1. What are the essential elements to the CDC method? Which publications describe what?

2. What did they do in their simulation study?

3. What should we do in ours?

Prejean 2011 PLoS One

The Point

Update to the 2008 JAMA paper, which provided 2006 incidence. This is 2007-2009 incidence, with "improved methodology"

Improved methodology

- Non-uniform, but rather data-driven, Pr(HIV test | within BED window)
- Missing transmission category - improved methodology (ref 10)

Of Note

- 34% increase in incidence among MSM in 2007-2009 compared to 2006 in ages 13-29; stable in other groups
- No backcalculation, just "stratified extrapolation method"

Hall 2013 JAMA Int Med

Backcalculation of overall and undiagnosed HIV prevalence in 4 steps:

1. Adjustment for reporting delay, incomplete reporting of diagnoses, missing transmission category

2. Backcalculation of cumulative # of those >13 infected by 2009

3. Overall prevalence = infected - deceased

4. Undiagnosed prevalence = overall prevalence - diagnoses

Our Simulation

Simulating true incidence

- Simple, like theirs? Baseline + trend, no trend, [illegible] trend? Trend over years, or over quarters?
- Time of infection is a random exact date within a quarter?

Simulating testing histories

How should the last negative test and 1st positive test bracket incidence?

- Use CDC data on recent infections for the 1st year since infection for some proportion of repeat testers, like in Prejean 2011?
- Do it differently for ever vs never tested, like they did?
- What would demonstrate the greatest impact on the method?

Measures to track

- Undiagnosed counts, not incidence?
- 95% CI coverage vs. coverage - we don't have 95% CI

Their simulation

Simulating true incidence

- 100 incident cases this interval with 15 yrs history (Poisson)
- Dates: seroconversion on EIA assay (uniform), AIDS (gamma), HIV testing (fixed, uniform or exponential)

Main features investigated

- Repeat testers: fixed (5 yrs ago) vs random first test date
- First testers: mean time to 1st test
- Both: trend in incidence (-10% per year, 0, +10% per year)

Measures tracked

- Mean relative bias in recovering the 100 infections in the year of interest
- CV
- 95% CI coverage

Karon 2008 Stat in Med

The Point

Methodology for estimating incidence from STAHRS data, "serologic testing algorithm for recent HIV seroconversion"

Data

HIV and AIDS diagnoses, ever tested Y/N/Miss, time since last neg?, recent infection result

Method

Extends incidence = prev/duration to situation where there is selection into BED testing.

Conclusions

Estimators are less biased than # diagnoses...
MI is important

Hall 2008 JAMA

The Point

Estimate STARHS incidence for 2006, via "stratified extrapolation method", and corroborate results with back-calculated incidence 1977-2006

Of Note

Appendix has a section on "extended back-calculation"

Overview

1. What are the essential elements to the CDC method? Which publications describe what?

2. What did they do in their simulation study?

3. What should we do in ours?

Hall 2008
JAMA

The Point

Estimate STARHS incidence for 2006, via "stratified extrapolation method", and corroborate results with back-calculated incidence 1977-2006

Of Note

Appendix has a section on "extended back-calculation"

Karon 2008
Stat in Med

The Point

Methodology for estimating incidence from STAHRS data, "serologic testing algorithm for recent HIV seroconversion"

Data

HIV and AIDS diagnoses, ever tested Y/N/ Miss, time since last neg?, recent infection result

Method

Extends incidence = prev/duration to situation where there is selection into BED testing.

Conclusions

Estimators are less biased than # diagnoses...
MI is important

"observed distribution of the time since the last negative HIV test for 219 persons who were BED recent and reported a last negative test date (median, 12 months; first and third quartiles, 5 and 20 months, respectively)"

Prejean 2011 PLoS One

The Point

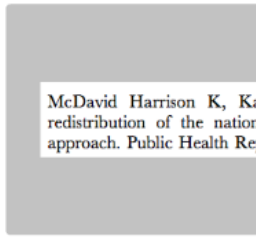
Update to the 2008 JAMA paper, which provided 2006 incidence. This is 2007-2009 incidence, with "improved methodology"

Improved methodology

- Non-uniform, but rather data-driven, Pr(HIV test | within BED window)
- Missing transmission category - improved methodology (ref 10)

Of Note

- 34% increase in incidence among MSM in 2007-2009 compared to 2006 in ages 13-29; stable in other groups
- No backcalculation, just "stratified extrapolation method"


McDavid Harrison K, Ka
redistribution of the nation
approach. Public Health Re

McDavid Harrison K, Kajese T, Hall HI, Song R (2008) Risk factor redistribution of the national HIV/AIDS surveillance data: An alternative approach. Public Health Rep 123(5): 618–627.

Hall 2013
JAMA Int Med

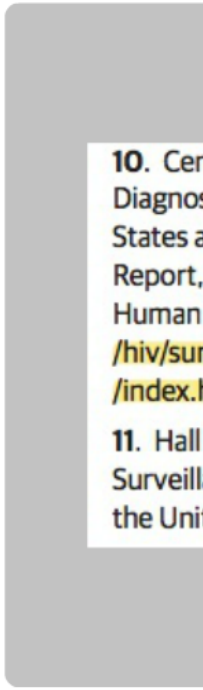
Backcalculation of overall and undiagnosed HIV prevalence in 4 steps:

1. Adjustment for reporting delay, incomplete reporting of diagnoses, missing transmission category

2. Backcalculation of cumulative # of those >13 infected by 2009

3. Overall prevalence = infected - deceased

4. Undiagnosed prevalence = overall prevalence - diagnoses



10. Centers for Disease Control and Prevention. Diagnoses of HIV infection and AIDS in the United States and dependent areas, 2010. HIV Surveillance Report, vol. 22. Atlanta, GA: US Dept of Health and Human Services, CDC; 2012. http://www.cdc.gov/hiv/surveillance/resources/reports/2010report/index.htm. Accessed May 25, 2012.

11. Hall HI, Song R, Rhodes P, et al; HIV Incidence Surveillance Group. Estimation of HIV incidence in the United States. *JAMA*. 2008;300(5):520-529.

Karon 2008
Stat in Med

The Point

Methodology for estimating incidence from STAHRS data, "serologic testing algorithm for recent HIV seroconversion"

Data

HIV and AIDS diagnoses, ever tested Y/N/ Miss, time since last neg?, recent infection result

Method

Extends incidence = prev/duration to situation where there is selection into BED testing.

Conclusions

Estimators are less biased than # diagnoses...
MI is important

Their simulation

Simulating true incidence

- 100 incident cases this interval with 15 yrs history (Poisson)
- Dates: seroconversion on EIA assay (uniform), AIDS (gamma), HIV testing (fixed, uniform or exponential)

Main features investigated

- Repeat testers: fixed (5 yrs ago) vs random first test date
- First testers: mean time to 1st test
- Both: trend in incidence (-10% per year, 0, +10% per year)

Measures tracked

- Mean relative bias in recovering the 100 infections in the year of interest
- CV
- 95% CI coverage

A. 2008;300(5):520-529.

Our Simulation

Simulating true incidence

- Simple, like theirs? Small + trend, no trend, small - trend? Trend over years, or over quarters?
- Time of infection is a random exact date within a quarter?

Simulating testing histories

How should the last negative test and 1st positive test bracket incidence?

- Use CDC data on recent infections for the 1st year since infection for some proportion of repeat testers, like in Prejean 2011?
- Do it differently for ever vs never tested, like they did?
- What would demonstrate the greatest impact on the method?

Measures to track

- Undiagnosed counts, not incidence?
- 95% CI coverage vs coverage - we don't have 95% CI

- Methods development
  - Individual rather than aggregate
    - Stochastic: simulate a time of infection
  - Compare uncertainty due to:
    - Size of dataset
    - Size of denominator population - undiagnosed *fraction*
    - Size/number/use of subgroups to get a total estimate?
    - Time of infection assumptions
      - Matt's midpoint of 16 and current age, rather than 16
    - % No
    - MAR vs correlated missing data
    - % missing data
    - Imputation of missing data/No - assumption vs multiple imputation?
    - Time step? Is quarterly data typical?
    - Stationarity of TID (check TID by year)
    - Self-report bias?
  - Use additional data for precision
    - Date of previous positive to shorten TID window
    - Incidence surveillance/risk-based testing for time of infection
    - Data from acute cases (Susan, 12/17 - ?)
  - NIR - redistribute to risk groups (CDC call)
  - Characteristics of those with missing data

CDC Paper Trail & Simulation

Overview

1. What are the essential elements to the CDC method? Which publications describe what?

2. What did they do in their simulation study?

3. What should we do in ours?

Prejean 2011 PLoS One

The Point

Update to the 2008 JAMA paper, which provided 2006 incidence. This is 2007-2009 incidence, with "improved methodology"

Improved methodology

- Non-uniform, but rather data-driven, Pr(HIV test | within BED window)
- Missing transmission category - improved methodology (ref 10)

Of Note

- 34% increase in incidence among MSM in 2007-2009 compared to 2006 in ages 13-29; stable in other groups
- No backcalculation, just "stratified extrapolation method"

Hall 2013 JAMA Int Med

Backcalculation of overall and undiagnosed HIV prevalence in 4 steps:

1. Adjustment for reporting delay, incomplete reporting of diagnoses, missing transmission category

2. Backcalculation of cumulative # of those >13 infected by 2009

3. Overall prevalence = infected - deceased

4. Undiagnosed prevalence = overall prevalence - diagnoses

Our Simulation

Simulating true incidence

- Simple, like theirs? (incid + trend, no trend, incid - trend)? Trend over years, or over quarters?
- Time of infection is a random exact date within a quarter?

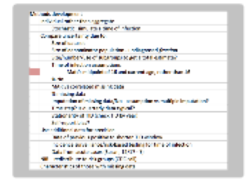
Simulating testing histories

How should the last negative test and 1st positive test bracket incidence?

- Use CDC data on recent infections for the 1st year since infection for (some proportion of repeat testers), like in Prejean 2011?
- Do it differently for ever vs never tested, like they did?
- What would demonstrate the greatest impact on the method?

Measures to track

- Undiagnosed counts, not incidence?
- 95% CI coverage vs. coverage - we don't have 95% CI



Their simulation

Simulating true incidence

- 100 incident cases this interval with 15 yrs history (Poisson)
- Dates: seroconversion on EIA assay (uniform), AIDS (gamma), HIV testing (fixed, uniform or exponential)

Main features investigated

- Repeat testers: fixed (5 yrs ago) vs random first test date
- First testers: mean time to 1st test
- Both: trend in incidence (-10% per year, 0, +10% per year)

Measures tracked

- Mean relative bias in recovering the 100 infections in the year of interest
- CV
- 95% CI coverage

Karon 2008 Stat in Med

The Point

Methodology for estimating incidence from STAHRS data, "serologic testing algorithm for recent HIV seroconversion"

Data

HIV and AIDS diagnoses, ever tested Y/N/Miss, time since last neg?, recent infection result

Method

Extends incidence = prev/duration to situation where there is selection into BED testing.

Conclusions

Estimators are less biased than # diagnoses...
MI is important

Hall 2008 JAMA

The Point

Estimate STARHS incidence for 2006, via "stratified extrapolation method", and corroborate results with back-calculated incidence 1977-2006

Of Note

Appendix has a section on "extended back-calculation"