MEASURING THE UNDIAGNOSED FRACTION:

Understanding the UW and CDC back-calculation models

Martina Morris, PhD  Director, UW CFAR SPRC
Jeanette K Birnbaum, PhD  Research Scientist, UW CFAR

Based on work originally developed by Ian Fellows, PhD
Outline

1. Back-calculation, the basics
   • The original method, developed in the 1980s

2. Current back calculation methods
   • UW: “Testing history” back-calculation
   • CDC: “Extended” back-calculation

3. Comparing the results for WA State
   • Future work
BACK-CALCULATION

The basics
Basic idea

• What you see now
• Is based on infections that happened in the past

Can you use new diagnoses to back-calculate past incidence?
Time from Incidence to Diagnosis

- Imagine an HIV Dx always happens within 3 years of infection
  - 25% get Dx in first year
  - 50% in second year
  - 25% in third year

distribution of “Time from Infection to Diagnosis”
Time from Incidence to Diagnosis

- With this TID (25%, 50%, 25%)
- And 100 new infections this year
- The *observed* HIV Dx curve in the future would look like this:

![Observed Diagnoses](chart)

**Observed Diagnoses**

- True incidence: 100 cases in year 1 only
- One year of incidence (n=100), distributed over 3 years of Dx

<table>
<thead>
<tr>
<th>Dx=1</th>
<th>Dx=2</th>
<th>Dx=3</th>
</tr>
</thead>
<tbody>
<tr>
<td>25%</td>
<td>50%</td>
<td>25%</td>
</tr>
</tbody>
</table>
# Tracking in tabular form

- From **infections** (unobserved) to **diagnoses** (observed)

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th># New Infections</th>
<th>Year of Diagnosis (t+z)</th>
<th>2013 (z=0)</th>
<th>2014 (z=1)</th>
<th>2015 (z=2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>100</td>
<td></td>
<td>25</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Obs Dx(t+z) = New Infections(t) * TID(t+z)**
With multiple years of incidence?

• Assume constant incidence (100 cases each year)
• And the same TID (25-50-25%)

• Annual observed HIV Dx is now a mix of cases from previous (up to 3) years

Diagnosis by year: Constant Incidence

Note: With constant incidence
Obs Dx = Incidence after max TID
With multiple years of incidence?

- Assume constant incidence (100 cases each year)
- And the same TID (25-50-25%)
- Annual observed HIV Dx is now a mix of cases from previous (up to 3) years

**Note:**

With constant incidence

\[
\text{Obs Dx} = \text{Incidence after max TID}
\]
With multiple years of incidence?

• Assume constant incidence (100 cases each year)
• And the same TID (25-50-25%)

• Annual observed HIV Dx is now a mix of cases from previous (up to 3) years

Diagnosis by year: Constant Incidence

Note:
With constant incidence and TID
Obs Dx = Incidence
after max TID
# Multi-year incidence, tabular form

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th>New Infections</th>
<th>Year of Diagnosis (t+z)</th>
<th>Future years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2013</td>
<td>2014</td>
</tr>
<tr>
<td>2013</td>
<td>100</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>2014</td>
<td>100</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>2015</td>
<td>100</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td><strong>Total Dx</strong></td>
<td><strong>300</strong></td>
<td><strong>25</strong></td>
<td><strong>75</strong></td>
</tr>
</tbody>
</table>

We’re assuming constant incidence

\[
\text{Obs Dx}(t+z) = \sum_{z=0}^{Z} \text{New Infections} \times \text{TID}(t+z)
\]

Future years

- TID t+0 = 25%
- t+1 = 50%
- t+2 = 25%
Undiagnosed cases calculation

<table>
<thead>
<tr>
<th>Year of Incidence</th>
<th>New Infections</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>2014</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>100</td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Totals</td>
<td>300</td>
<td>25</td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>

Undiagnosed (2015) = Cumulative incidence – Cumulative diagnosed

= 300 – 200

= 100
This would be straightforward, ... *if*

• If you could observe everything
  • Incidence
  • TID
  • Dx cases

• But we only observe Dx cases...

• *If we can estimate the TID from some other data*, then we can “back calculate” the new infections, and the undiagnosed cases

\[
\text{Obs \: Dx}(t+Z) = \sum_{z=0}^{Z} \text{New Infections} \ast \text{TID}(t+z)
\]
Start with observed Dx

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Dx</strong></td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Obs Dx( t+z ) = \sum_{z=0}^{Z} \text{New Infections} * \text{TID}(t+z)
Use an estimated TID

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>25</td>
<td>50</td>
<td>0.25*NI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>25</td>
<td></td>
<td>0.50*NI</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td>0.25*NI</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total Dx</strong></td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{Obs Dx}( t+z ) = \sum_{z=0}^{Z} \text{New Infections} \times TID(t+z)
\]
Back-fill in all the cells

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>0.25*NI</td>
<td>0.50*NI</td>
<td>0.25*NI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>0.25*NI</td>
<td>0.50*NI</td>
<td>0.25*NI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>0.25*NI</td>
<td>0.50*NI</td>
<td>0.25*NI</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Dx</strong></td>
<td><strong>25</strong></td>
<td><strong>75</strong></td>
<td><strong>100</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Obs Dx( t+z ) = \sum_{k=0}^{Z} \text{New Infections} * \text{TID}(t+z)
And solve for the NI (New Infections)

<table>
<thead>
<tr>
<th>Year of Diagnosis (t+z)</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of Incidence (t)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>2015</td>
<td>100</td>
<td></td>
<td>25</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>Total Dx</td>
<td>300</td>
<td>25</td>
<td>75</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

\[
\text{Obs Dx}(t+z) = \sum_{z=0}^{Z} \text{New Infections} \times \text{TID}(t+z)
\]
Note (1)

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th>New Infections</th>
<th>Year of Diagnosis (t+z)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2013</td>
</tr>
<tr>
<td>2013</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>2014</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>2015</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td><strong>Total Dx</strong></td>
<td><strong>300</strong></td>
<td><strong>25</strong></td>
</tr>
</tbody>
</table>

We assume constant incidence here, but the approach also works if incidence is changing.
# Note (2)

<table>
<thead>
<tr>
<th>Year of Incidence (t)</th>
<th>New Infections</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>100</td>
<td>25</td>
<td>50</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Dx</strong></td>
<td><strong>300</strong></td>
<td><strong>25</strong></td>
<td><strong>75</strong></td>
<td><strong>100</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We assume a constant TID here too, but the approach also works if the TID is changing.
We can also use this method to project diagnoses forward.
Summary

\[ \text{Observe} \]

\[ D_x(t+z) = \]

\[ \sum_{z=0}^{Z} \text{New Infections} \times TID(t+z) \]

\[ \text{Back calculate} \quad \text{Estimate from other data} \]
ORIGINAL BACK-CALCULATION

From the way way back
Context

- First used in 1986 by Brookmeyer and Gail for the HIV epidemic
  - At that time, only AIDS Dx were available
  - So the goal was to back calculate HIV incidence from AIDS Dx

Original back-calculation

**Observe**

\[ \text{AIDS \, \text{Dx}(t+z)} = \]

**Uses AIDS \, \text{Dx}, and AIDS \, \text{TID}**

**Back calculate**

\[ \sum_{z=0}^{Z} \text{HIV Incidence}(t+Z-z) \times \text{AIDS \, \text{TID}(t+z)} \]

**Estimate from data**

*Data sources (all from the 1980s):*
- Multicenter Hemophilia Cohort Study (N=373)
- International Registry of Seroconverters (MSM, N=1020)
- Amsterdam cohort studies (IDU, N=173; MSM, N=348)
- Multicenter AIDS Cohort Study (MSM, N=1861)
Estimating the AIDS TID ("Incubation Period")

**Multiple approaches**
- Different parametric forms (Weibull, Gamma)
- Different # stages of infection (2-5)

**Problems**
- Time of infection usually not known
- Loss to follow-up
- Representativeness of cohorts
- Treatment delays AIDS onset
Key improvement to the method

Incorporating data on HIV Dx

- Necessary because treatment dramatically reduced AIDS Dx
- Possible because HIV Dx became reportable
Fast forward to now

Before testing + treatment

Original AIDS Back-Calc

Current

UW “Testing History” Back-Calc

CDC “Extended” Back-Calc
UW vs CDC: Overview

• Similarities:
  • Both incorporate data on observed HIV Dx
    • But use this in very different ways
  • Both use complex algorithms for estimation
    • UW : EM Algorithm (EM = Expectation-Maximization)
    • CDC : Bayesian MCMC (MCMC = Markov Chain Monte Carlo )

• Differences:
  • Use different information in the back-calc process
  • Select and weight cases from the Dx populations differently
UW TESTING HISTORY METHOD

Originally developed by Ian Fellows, PhD
UW Testing History back-calculation

\[ \text{HIV Dx}(t+z) = \]

A simple approach: Use HIV Dx to back calculate HIV incidence

\[ \sum_{k=0}^{Z} \text{HIV Incidence}(t+Z-z) \times \text{HIV TID}(t+z) \]

So for this we need to estimate the HIV TID
Use testing histories to estimate the TID

• Testing histories give us an infection window, but...

  • When did infection occur in the window?

  • What if people never had a LNT, or have missing data?
Full Model overview (we’ll take it in pieces)

Stratify by the LNT

1. Last Negative Test (LNT)
   - Yes: Window = ITI
   - No: Window = min(18, age-16)
   - Missing: Assume MAR

TID Estimation:
- Base Case: Uniform distn across window
- Upper Bound: All inf at start of window

3 important assumptions

 Undiagnosed estimation

Incidence Back-Calculation
1. Model for repeat testers

**Stratify by the LNT**

- **Yes**
  - Window = ITI
  - \( ITI = \text{inter-test interval} \)

- **No**

- **Missing**

**TID Estimation:**

- **Base Case**
  - Uniform distn across window

- **Upper Bound**
  - All inf at start of window

**Undiagnosed estimation**

**Incidence Back-Calculation**

**First assumption:**
Distribution of infection probability
**Assumption 1:**

**Infection probability distribution**

**Base Case**

Uniform:
Distributes the probability of infection uniformly across the possible interval

**Upper Bound**

At last neg test:
Probability = 1 that infection occurred on the day after the last negative test
2. Model for Dx with no previous test

Stratify by the LNT

- Last Negative Test (LNT)
  - Yes
  - No
  - Missing

Second assumption: Window length

Window = \( \min(18, \text{age}-16) \)

TID Estimation:
- Base Case: Uniform distn across window
- Upper Bound: All inf at start of window

Undiagnosed estimation

Incidence Back-Calculation
**Assumption 2:**

**Window length if no previous test**

- 95% of HIV+ progress to AIDS in 18 years (Lui 1996)
- Age 16 is the median age of sexual debut in the US
- So we take the minimum of these as the window length
- And then apply base case or upper bound assumption for the distribution of infection probability

![Graph showing the window of possible infection](image)
3. Model for cases missing test info

Stratify by the LNT

Last Negative Test (LNT)

- Yes
- No
- Missing

TID Estimation:

- Base Case
  - Uniform distn across window
- Upper Bound
  - All inf at start of window

Undiagnosed estimation

Incidence Back-Calculation

Assume MAR

Third Assumption: Treatment of missing cases
Assumption 3: If Dx case is missing test information

• We have two options:

  • **Include** these when estimating the TID distribution
    
    • Assuming the maximum possible infection window
    • **IMPACT**: A conservative (longer) estimate of the time spent undiagnosed.

  • **Exclude** these when estimating the TID distribution
    
    • We still use them in the back-calculation, we just give them the TID estimated from the other cases
      • Assumes these cases are “missing at random” (MAR)
Full Model overview

Stratify by the LNT

Last Negative Test (LNT)

1. TID Estimation:
   - Base Case: Uniform distn across window
   - Upper Bound: All inf at start of window

2. Window = ITI
   - Yes
   - No: Window = min(18, age-16)

3. Assume MAR

3 important assumptions

Undiagnosed estimation

Incidence Back-Calculation

Friday, May 27, 2016
SPRC-PHSDK Lunchbox Talks
Results WA State: ITI dist’n (2006-2015)

• The distribution of inter-test intervals
  • For all non-missing cases

46% have an LNT (24% HIV/AIDS Dx)

12% no LNT (51% HIV/AIDS Dx)

42% of cases are missing, so not included in TID estimation (39% HIV/AIDS Dx)
Results WA State: HIV TID (2006-2014)

Test found no significant variation in TID over time

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Mean TID</th>
<th>% UnDx at 1yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Case</td>
<td>2.5 years</td>
<td>45%</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>5 years</td>
<td>64%</td>
</tr>
</tbody>
</table>
Results WA State: Incidence & UnDx

Observed Dx (black)

Estimated Incidence (colors)

Obs=Est and Base=UB suggests relative stability in the dynamics

Estimated UnDx cases

Upper bound ~ 2x Base Case

2014:
~100 new cases/qtr
1200-2500 UnDx cases
CDC Extended Back-Calc Overview

- Uses AIDS Dx, like original method
- Adds data on HIV Dx

- Stratifies observed cases by
  - HIV Dx only vs.
  - Concurrent HIV/AIDS diagnosis (AIDS Dx within 1 year of HIV Dx)

- And it uses a Bayesian estimation approach
  - So it will need “prior” distributions to start the estimation algorithm
CDC Model overview

Data Cleaning → Dx Cases

1. AIDS TID
2. “Priors” on testing rate distributions
3. “Priors” on incidence distributions

Stratifies by concurrent HIV/AIDS dx

AIDS in same year
HIV only

Undiagnosed estimation

Incidence AND Testing rate Back-Calculation

Priors are part of the “Bayesian” estimation method
Key difference: competing risks of Dx

• In the UW TH model
  • HIV+ person faces only one “risk”: the risk of HIV Dx by testing

• In the CDC model
  • HIV+ person faces two risks: AIDS Dx or HIV Dx
  • So the model has to specify how these processes unfold over time

<table>
<thead>
<tr>
<th>Event at time t</th>
<th>Depends on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS Dx</td>
<td>AIDS TID (Dx), not tested</td>
</tr>
<tr>
<td>HIV Dx</td>
<td>AIDS TID (noDx), not tested earlier, tested now</td>
</tr>
<tr>
<td>UnDx</td>
<td>AIDS TID (noDx), not tested</td>
</tr>
</tbody>
</table>
Assumption 1: Historical AIDS TID

This is how they distribute the probability of HIV infection back in time for someone diagnosed with AIDS.
Assumption 2: HIV testing rate priors

HIV testing rates are not observed – they are estimated

Huh?

But these rates are used to estimate incidence. So how can they be estimated too?
Bayesian estimation in a nutshell

• Say you have a parameter you want to estimate, $p$
  • But you don’t have a simple formula for it in terms of the data you observe (here HIV Dx and AIDS Dx)

• Draw a starting value from a distribution
  • Reflects what we think/know about the value of $p$

• Plug it in and calculate the predicted Dx
• Compare the predicted Dx to the observed
• Update the estimate of $p$ in the prior
• Repeat until predicted=observed

This is the prior distribution
Assumption 2: HIV testing rate priors

HIV testing rates are not observed – they are estimated

- Prior distributions for the annual rates, 1977-present

The testing rate is for HIV+ persons (not the pop’n)
Assumption 3: HIV incidence prior

HIV incidence is also estimated in this model

- Specifies another “prior distribution” to start

Single prior, but posterior estimates allowed to vary by year
Overview of the CDC model

- Two unknown parameter sets to estimate:
  - Annual testing rates \( \{p_i^H\} \)
  - Annual HIV incidence \( \{\lambda_i\} \)

- Two observed data series
  - Annual HIV Dx
  - Annual AIDS Dx

- One fixed assumption (the AIDS TID)

- Two Bayesian priors
  - for testing rates and HIV incidence
CDC Results: National testing rates 1985-2010

Mean rate estimate stabilizes at around 22% per year
CDC Results: National mean time to HIV Dx

2010: 3.3 years

Expected Time–Since–Infection in United States

(Note: WA analysis starts in 2006, est 2.5 - 5 yrs)
## CDC Results: National UnDx estimate for 2012

<table>
<thead>
<tr>
<th>CDC National Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Undiagnosed Fraction</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Undiagnosed</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Note:* This requires estimating the number of persons living with HIV, and we’re not describing how they do that here.
Other aspects of both models

• Both can be used to estimate the UnDx *Fraction*
  • But need an estimate of PLWH for the denominator

• Both models assume there is some stability in the year to year changes (smoothing)
  • For UW method: Incidence counts in adjacent years are smoothed
  • For CDC method: Both incidence and testing rates are smoothed

• Both models can stratify estimation by other factors
  • Sex
  • Risk exposure
  • Geography (*though this raises issues about modeling migration*)
  • *But small sample sizes will lead to unstable estimates*
Summary of model differences

• Uses of HIV Dx
  • UW estimates the HIV TID to back-calculate HIV incidence
    • Using measured inter-test intervals for HIV Dx when available
    • Relies on the max AIDS TID window for cases Dx on their first test
    • Variation in the TID by year can be evaluated and incorporated
  • CDC estimates testing rates as part of the HIV incidence back-calculation
    • Calibrated to best fit observed HIV and AIDS Dx trends
    • Relies also on the AIDS TID
    • The annual average rate is allowed to vary over time

• Uses of AIDS Dx
  • UW does not use this (but could be adapted)
  • CDC uses this to estimate both testing rates and HIV incidence
COMPARING RESULTS FOR WA

What we know now, and plans for future investigation
## WA state estimates for 2012

<table>
<thead>
<tr>
<th></th>
<th>CDC*</th>
<th>UW</th>
<th>% DIFFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UnDx Fraction</strong></td>
<td>11.0</td>
<td>10.6</td>
<td>-3.8%</td>
</tr>
<tr>
<td></td>
<td>(7.7—15.0)</td>
<td>(18.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Undiagnosed</strong></td>
<td>1,700</td>
<td>1,410</td>
<td>-20.6%</td>
</tr>
<tr>
<td></td>
<td>(1,200--2,400)</td>
<td>(2,750)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>15,500</td>
<td>13,310</td>
<td>-16.5%</td>
</tr>
<tr>
<td></td>
<td>(14,900--16,100)</td>
<td>(14,650)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CDC*</th>
<th>UW</th>
<th>% DIFFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSM</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UnDx Fraction</strong></td>
<td>11.7</td>
<td>6.8</td>
<td>-72.1%</td>
</tr>
<tr>
<td></td>
<td>(7.5—16.5)</td>
<td>(12.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Undiagnosed</strong></td>
<td>1,200</td>
<td>647</td>
<td>-85.5%</td>
</tr>
<tr>
<td></td>
<td>(730--1,800)</td>
<td>(1274)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>10,300</td>
<td>9,519</td>
<td>-8.2%</td>
</tr>
<tr>
<td></td>
<td>(9,900--10,800)</td>
<td>(10,147)</td>
<td></td>
</tr>
</tbody>
</table>

*Source: WA DOH*
Potential Sources of Differences

• PLWH denominator for estimating the undiagnosed fraction
  • Doesn’t seem to be the driving difference since the totals are similar

So, that leaves one of:

• Case selection/weighting
  • CDC does adjustments and weighting for reporting delays and missing data
  • But I’m guessing this is not the primary driver

• Model structure and assumptions
  • If testing histories provide more precision for MSM, our estimates may be better
  • Not sure what could lead to their estimates being better
Future Work

- Compare results using identical datasets
  - Accommodate weights in testing history method to use CDC data
  - Can’t run CDC method on our data since it needs to go back to 1977, which requires doing all their data cleaning relevant to the older cases

- Test both models on a mock dataset in which incidence is known
  - “Simulation study”
  - Generate the mock data from an independent model of HIV natural history
THANK YOU