Estimating the Undiagnosed Fraction: A new "Testing History" approach

Martina Morris
presenting joint work with
Ian Fellows, PhD (statistics) and
Matt Golden, MD

Motivation

- Currently have 2 national estimates of the undiagnosed fraction
  - Empirical estimate from NHBS for MSM: **44%**

    Chen M, Rhodes PH, Hall IH, Kilmarx PH, Branson BM, Valleroy LA, Centers for Disease C, Prevention: **Prevalence of undiagnosed HIV infection among persons aged >/=13 years--National HIV Surveillance System, United States, 2005-2008.** ***MMWR Morb Mortal Wkly Rep 2012, 61 Suppl:57-64.***
    - Seattle NHBS MSM estimate is ~ 15%
  - Back-calculation estimate from CDC National HIV surveillance system: **19.1%**

    Hall HI, Frazier EL, Rhodes P, Holtgrave DR, Furlow-Parmley C, Tang T, Gray KM, Cohen SM, Mermin J, Skarbinski J: **Differences in human immunodeficiency virus care and treatment among subpopulations in the United States.** ***JAMA internal medicine 2013, 173(14):1337-1344.***

Questions about these estimates

- NHBS MSM estimate
  - Sample is venue based. Representative?
  - Status awareness is self-report. Non-disclosure?
  - Local/regional differences from national patterns?

- CDC national back-calculation estimate
  - Method is not well described
  - Depends on assumptions about the distribution of time from infection to AIDS
  - Appears to make many other assumptions as well

Goal: A tool for local use

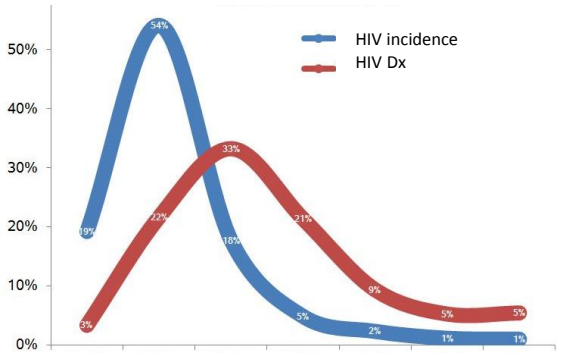
Method based on testing history data only

1. Back-calculation based approach
   - Uses HIV Dx + most recent HIV negative test
   - Can provide software (R) to local Public Health Depts
   - Will focus on this in the current presentation

2. Back of the envelope calculation
   - Based on Matt Golden's idea
   - How bad can it be? ☺
   - Turns out it perfectly matches the back-calculation estimate (under some assumptions)
   - Can provide excel worksheet to local Public Health Depts

Back calculation: the idiot's guide

- Basic idea
  - What you see now
  - Is based on infections that happened in the past
  - So: can you use new diagnoses to back-calculate what happened in the past?



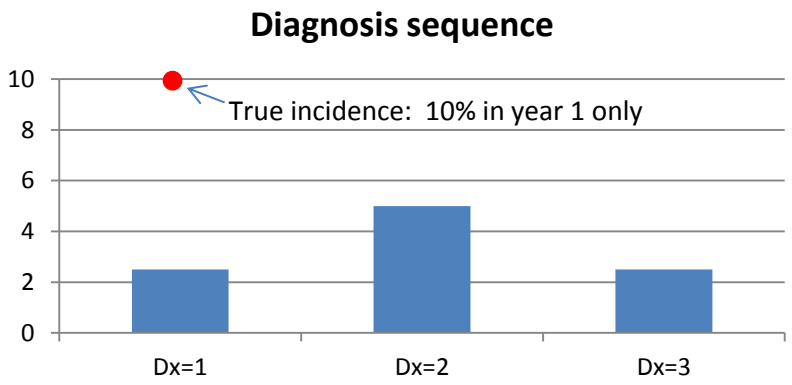
Imagine smearing the incidence of new HIV infections forward as they are diagnosed over time

Some are diagnosed quickly, other not

What we see now = sum over time of incidence at time t * probability(Dx at time t+Z)

Breaking it down

- Imagine a Dx always happens within 3 years of HIV infection
  - 25% get Dx at t=1
  - 50% at t=2
  - 25% at t=3
- If there was only one year of 10% incidence, the observed HIV Dx curve would look like this:

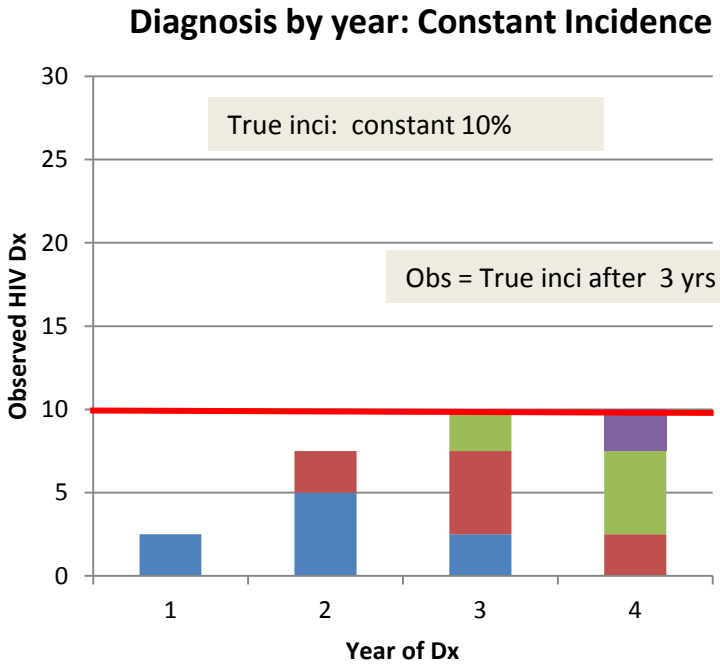


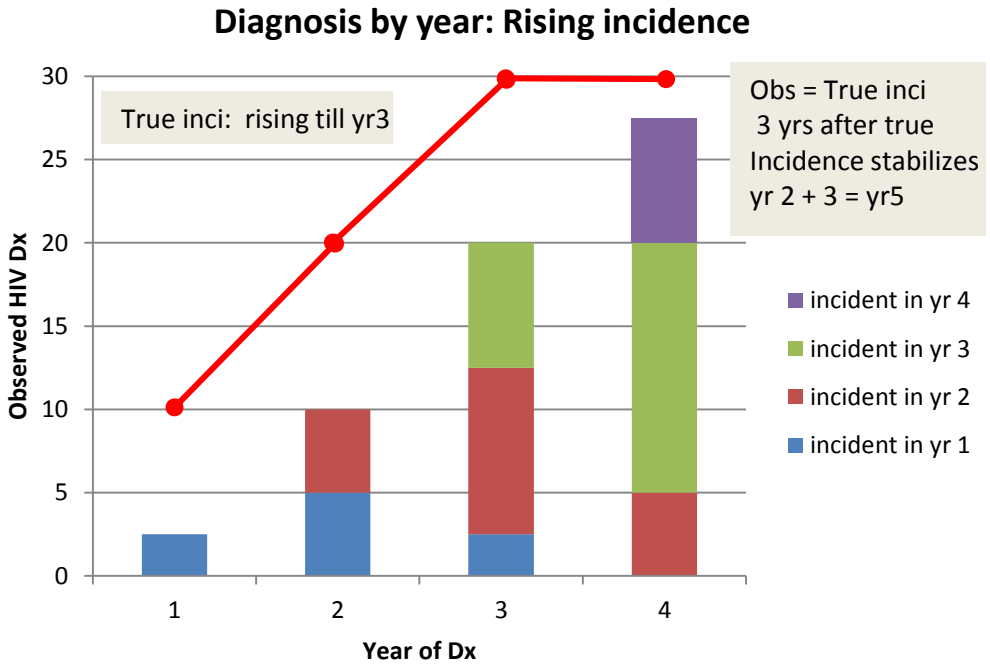
With only one year of incidence, smeared over 3 years of Dx

For any single year:
Observed Dx ≠ True Incidence

With multiple years of incidence?

- Assume the same Dx rates: 25-50-25%
- You only observe the total HIV Dx, which is now a mix of cases from previous (up to 3) years





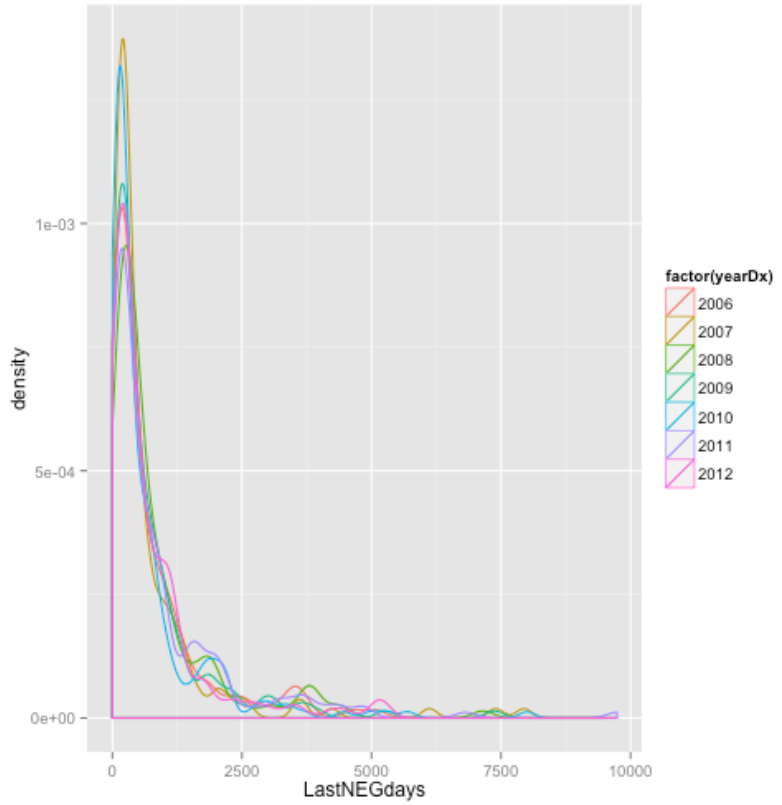
Obs Dx(t+Z) = sum of (HIV incidence at time t * probability(Dx at time t+Z)

Variants of back-calculation

- Standard form
  - Estimate HIV incidence from AIDS Dx
  - Assumptions: time from infection to AIDS Dx
- "Extended" form
  - Estimate HIV incidence using both HIV Dx and AIDS Dx
    - Can also include biomarkers, e.g., CD4, recent infection
  - Assumptions: time from infection to AIDS Dx, time from infection to HIV Dx, impact of symptoms on testing rate
- Our version: "Testing History" method
  - Estimate undiagnosed fraction from HIV test dates: last – to +
  - Assumption: time from infection to HIV Dx

Basic idea

Testing history data:
Time from last negative test to Dx



- For HIV Dx with a previous negative test
- We know infection must have happened in this interval
- If we assume a distribution for the probability of infection in that interval
  - we can estimate incidence
  - and the undiagnosed fraction
  - we'll examine 2 different assumptions

Testing history back-calculation

- Start by back-calculating incidence from the convolution:

$$Dx(t) = \sum_{s<t} Ni(s) * p(Dx = t \mid inf = s)$$

Diagnosed cases at time t (known)

Number of incident * cases at time s (unknown)

Probability of diagnosis at time t given infected at time s (assumed, based on test history)

- Then use incidence to estimate the undiagnosed fraction

$$Ux(t) = \sum_{s<t} Ni(s) * p(Dx > t \mid inf = s)$$

Sensitivity to assumptions

Assumptions can influence estimates, so we explore different assumptions

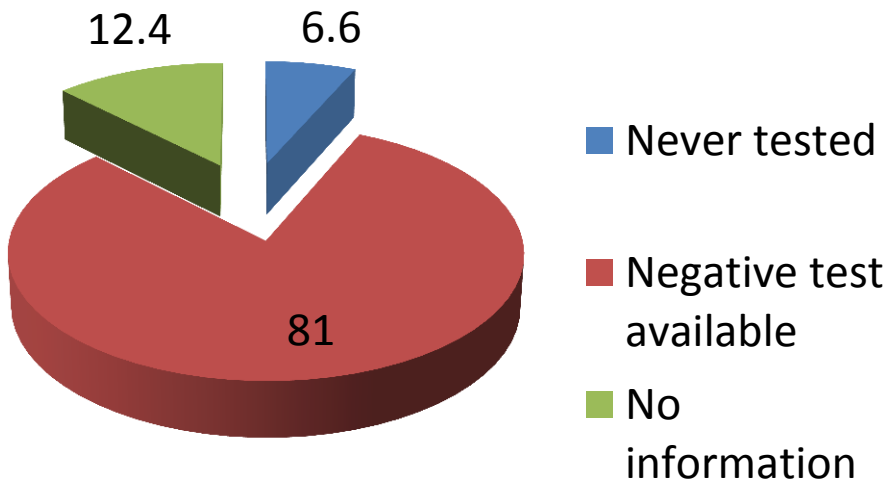
- *First: Time from Infection to HIV Dx (the TID distribution)*
  - **Base case:** assume uniform distribution of infection across interval
  - **Upper bound:** assume infection immediately after last negative test.
- *Second: Incidence change over time*
  - **Time varying** (each year can be different)
  - **Constant** (the estimating equation then simplifies dramatically)

Application

- To all new MSM HIV Dx in King County
  - 85% of new Dx are in MSM in KC
  - Most MSM have a previous negative test
- Timeframe: 2006-2012
- 3 sources of data on testing history
  - eHARS: only includes validated test histories
  - HIS: CDC testing and Tx history Qx (self report to DIS)
  - PS: Partner services data (self report to DIS)

Testing Data

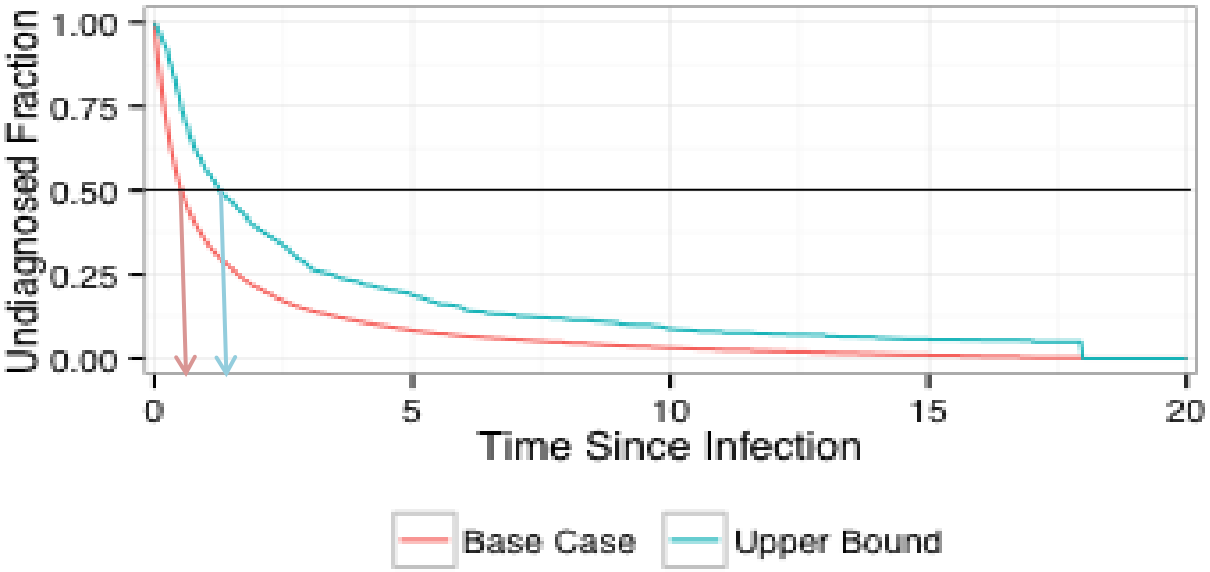
- New MSM diagnoses 2006-12: 1522
- Testing history:
  ~88% known
- Sourced from:
  - 25% eHARS
  - 71% HIS
  - 31% Partner Services



Correlations: (HIS,eHars) = .76; (HIS,PS) = 0.85

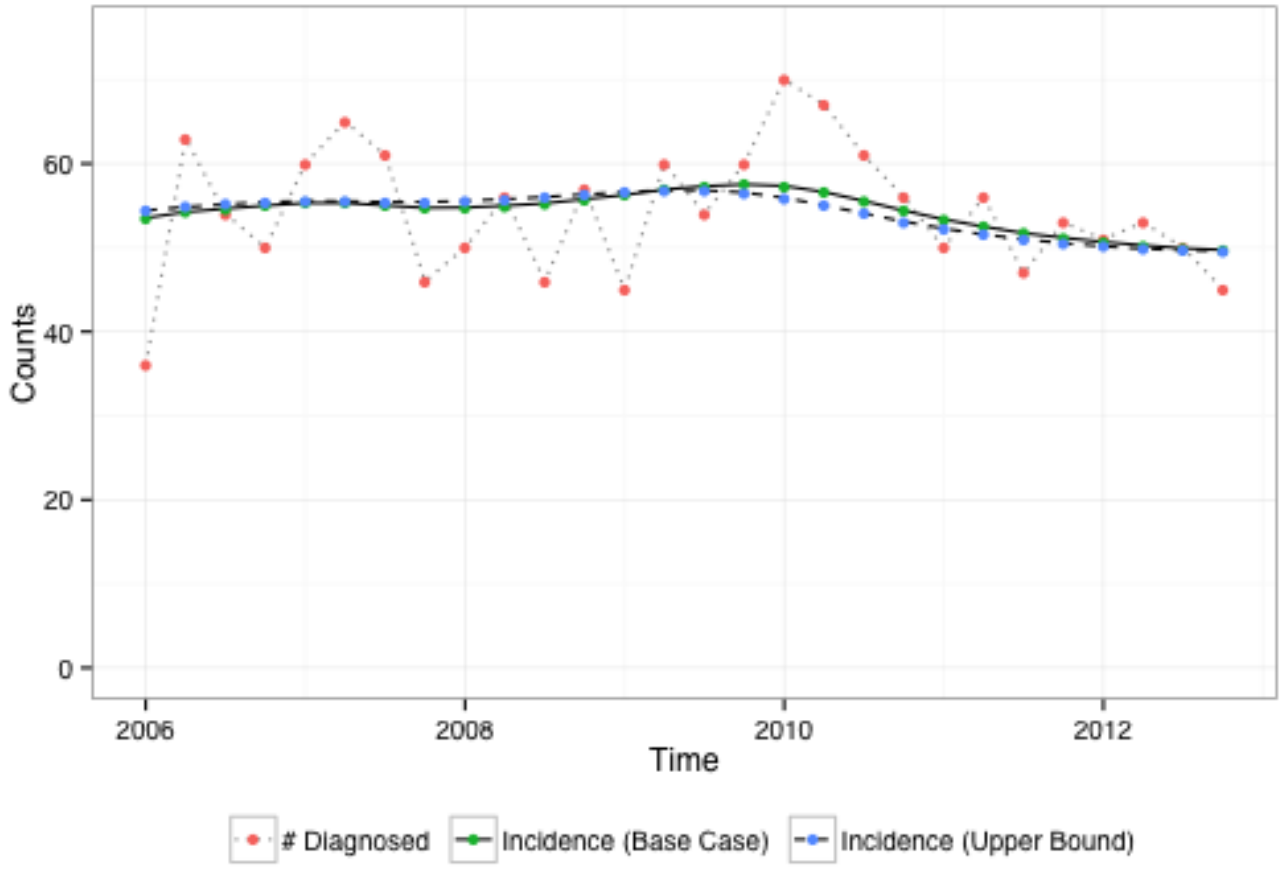
Estimates of TID (dist'n of time from infection to HIV Dx)

Median estimates of TID: Base case = 0.5 years
Upper bound = 1.3 years



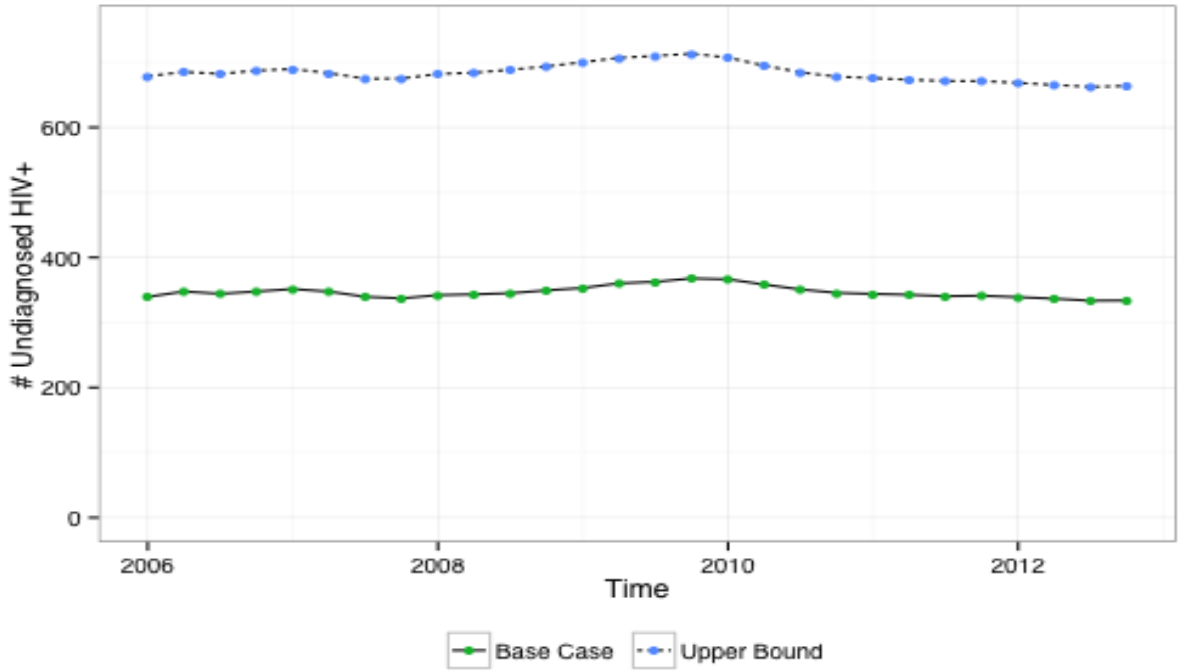
Empirical estimates of possible infection interval: Mean = 3.12 years
Median = 1.25 years

Incidence Count Estimates



No difference between base case and upper bound

Estimated Undiagnosed Cases (Count)



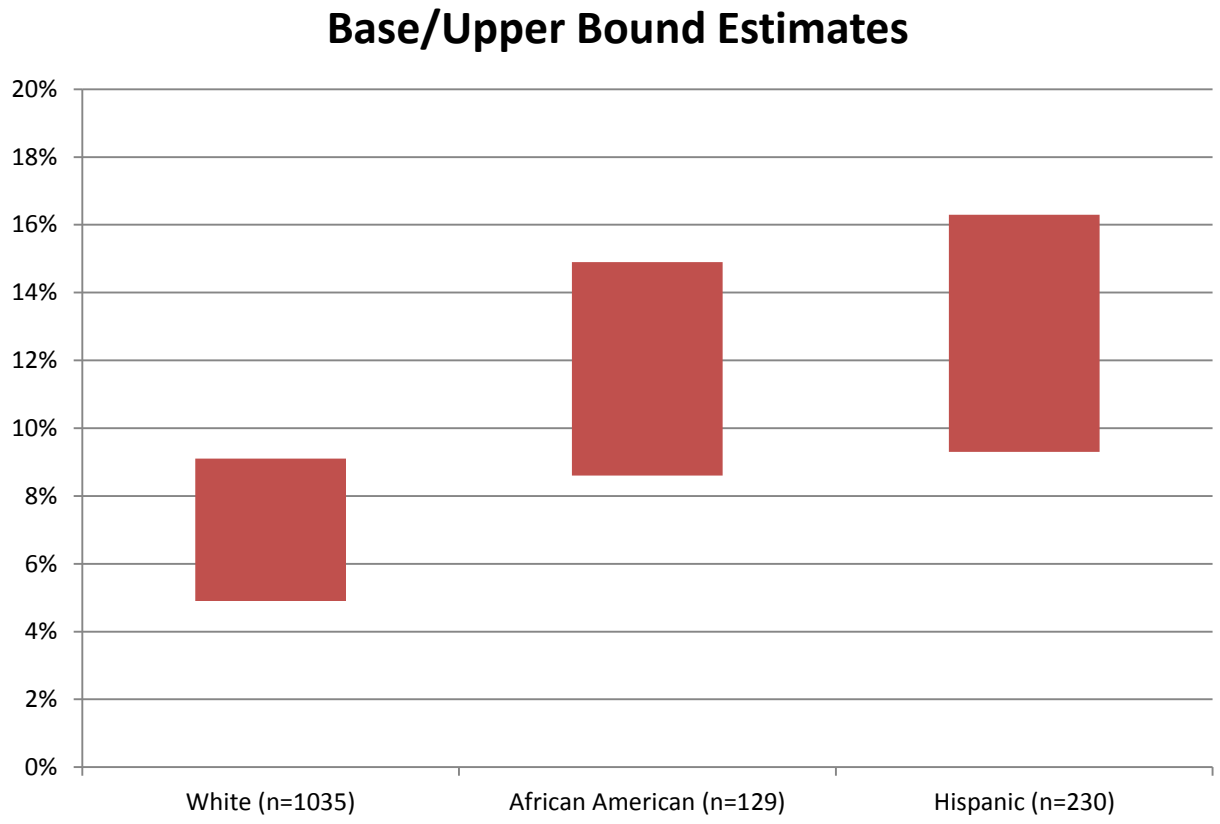
Upper bound estimate is ~double the base case

Undiagnosed fraction estimates

TID Distribution Assumption	Incidence Model	Estimated Number of MSM with HIV	Estimated Number of MSM with undiagnosed HIV	Undiagnosed Fraction
Base case	Varying	5850-5884	333-368	5.7%-6.3%
	Constant	5863	347	**5.9%**
Upper bound	Varying	6178-6229	662-713	10.7%-11.4%
	Constant	6203	687	**11.1%**

Assumptions matter, but both estimates are quite low: 5.9 – 11.1%

Undiagnosed fractions: by race/ethnicity



Key strengths of this approach

Compared to other back-calculation approaches:

- This approach does not use data on AIDS Dx
  - So we don't need to make assumptions about the time from infection to AIDS
- This approach uses the observed testing frequency
  - So we don't need to make assumptions about rate of testing, and whether it changes over time
  - And changes in test frequency will be accurately reflected by the estimate

Limitations

- Of the approach
  - Need robust testing history data : 81% have a last negative test date in the Seattle data
    - The sensitivity of this estimate to the TDI assumptions will increase as the number of cases with no prior test data rises.
    - One could potentially model this with missing data methods
  - We assume testing is not correlated to risk behavior
    - But if recent risk leads to testing, then the undiagnosed fraction is probably lower than the base case estimate
    - This, too, can be modeled

- Of the Seattle analysis
  - Only 25% of our cases have a chart-validated last negative test date
  - But the correlations of the self-report dates with eHARS suggest good validity for the rest