



# Estimating HIV Incidence from HIV Reporting Data

---

John Karon, CDC

Rick Song, CDC

Ed Kaplan, Yale

Ron Brookmeyer, Johns Hopkins

# How Many New HIV Infections Occur in the US Each Year?

---

- ◆ Nobody really knows (though CDC has suggested 40,000/yr for the last decade)
- ◆ Recent Institute of Medicine panel on HIV prevention strategy in the US recommended that CDC focus on HIV incidence
- ◆ CDC's own strategic plan recognizes need for estimating HIV incidence

# Why Bother?

---

- ◆ Rates of new HIV infection needed to:
  - Monitor epidemic
  - Evaluate interventions (and overall CDC plan, which calls for cutting rate of new infections in half!)
  - Allocate resources

# What Are The Challenges?

---

- ◆ Until recently, only newly diagnosed AIDS cases were regularly reported
  - Backcalculation, or reconstruction of HIV epidemic from AIDS reporting data, used to work, but not anymore (individual variation in HIV treatments plus success of treatments destroys statistical pattern of HIV→AIDS progression), and doesn't give recent incidence
- ◆ Cohort studies take too long and cost too much

# What's New? (Part I)

---

- ◆ Laboratory-based procedures for estimating HIV incidence from *single* samples (snapshot methods)
  - Snapshot theory worked out (Kaplan and Brookmeyer), examples with CD4 and CD4%
  - p24 (Brookmeyer and Quinn)
- ◆ Clear winner: detuned assay (Janssen, Satten, Busch *et al*)

# Detuned Assay Basics

(Parekh and McDougal, *AIDS Rev* 2001;3:183-193)

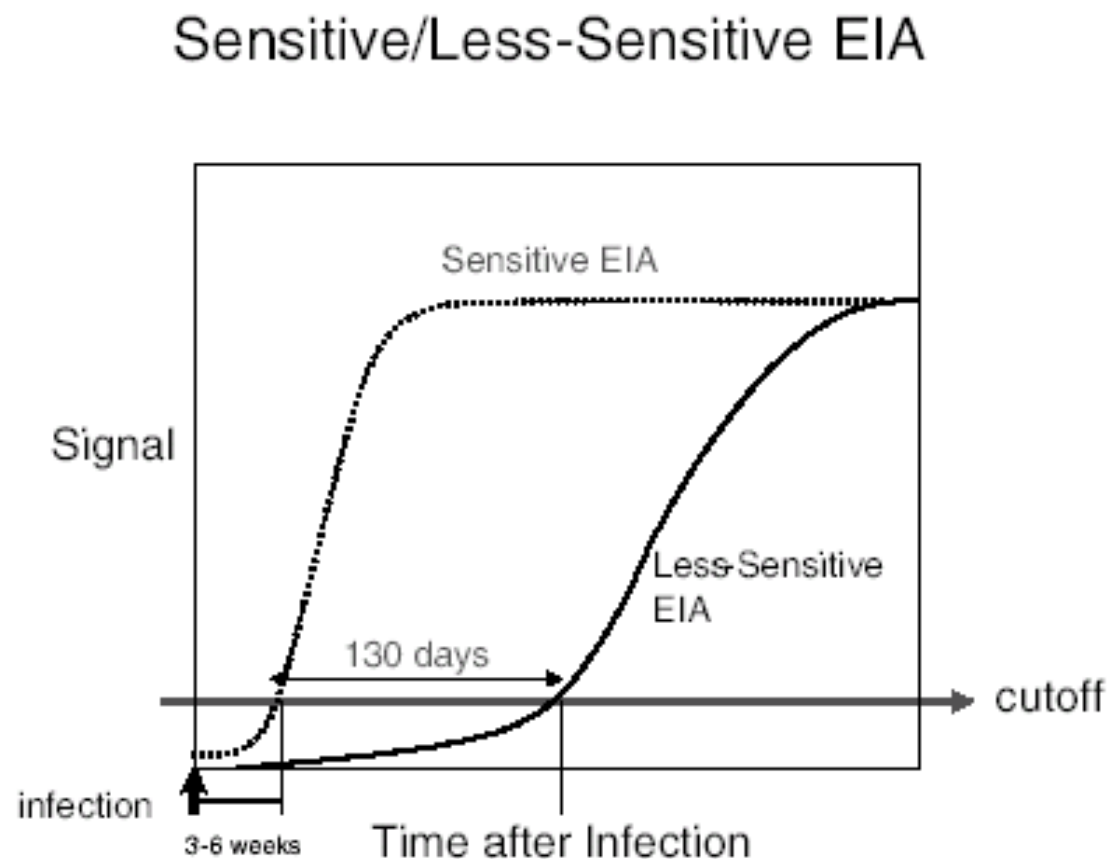
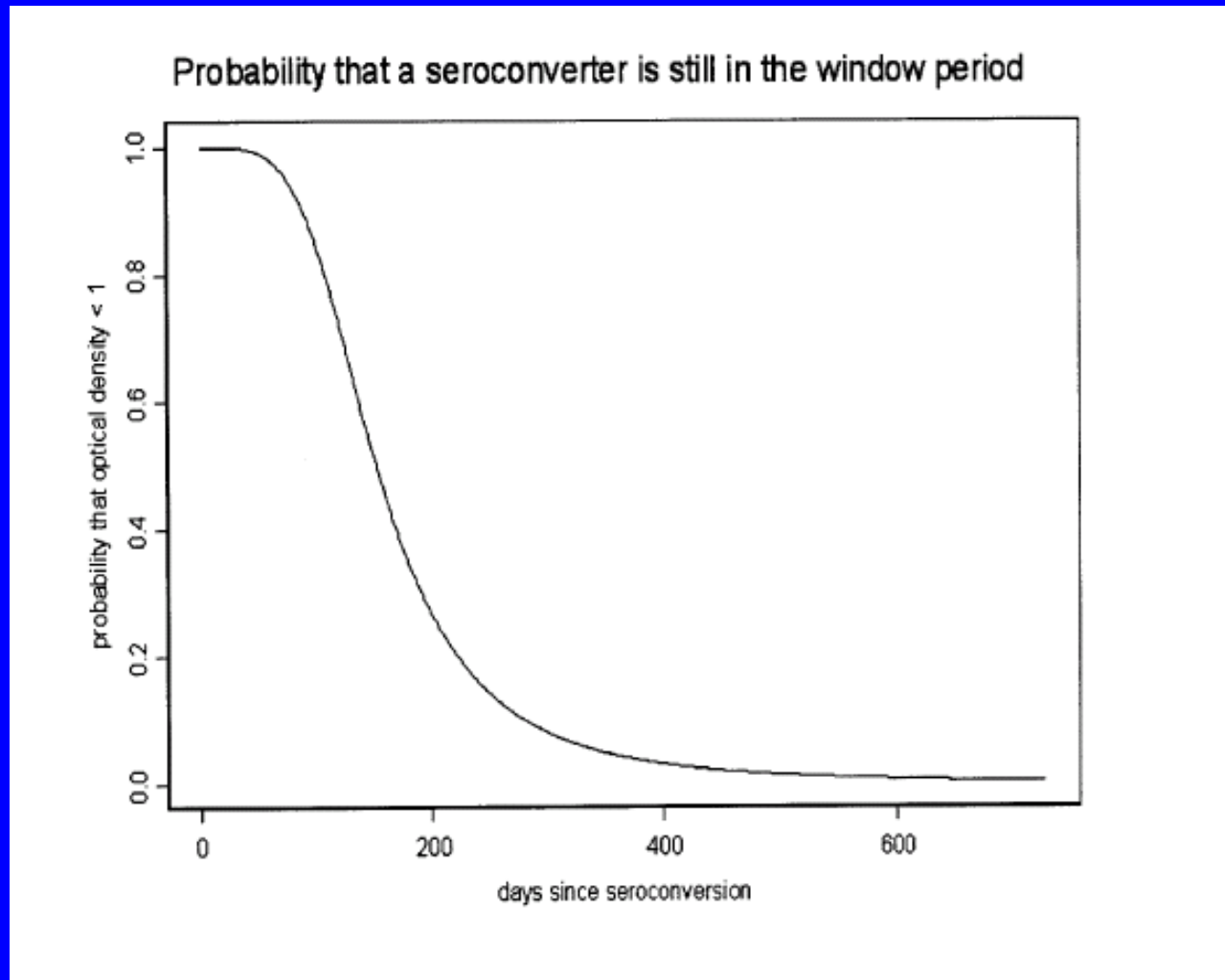


Figure 4. A: Comparison of sensitive (3A11) and less-sensitive (3A11-LS) EIAs. B: Sensitive EIA plateaus soon after seroconversion, while less-sensitive EIA has longer dynamic range and takes about an additional 130 days to register reactive.

# Detuned Assay Basics (Organon Teknika, Subtype B, courtesy Bob Byers)



# What's New? (Part II)

---

- ◆ HIV reporting is now standard policy in many states (34 as of 6/1/01)
- ◆ CDC expects most states to adopt HIV reporting

# Framework for Estimating HIV Incidence

---

- ◆ Every new HIV infection in the US will either:
  - be captured in an HIV reporting database
  - or not!
- ◆ Leads to following identity:

$$\# \text{ New Infections} = \# \text{ Reported} + \# \text{ Not Reported}$$

# Ideas For HIV Incidence Among Those Who Don't Test

---

- ◆ IOM proposal – sentinel sites: need to estimate incidence at sites (e.g. STD clinics, drug treatment programs, hospitals), and also chance of getting sampled
- ◆ Different idea – community surveys to estimate fraction that test; incidence in those that don't
- ◆ Different idea – focus on infected persons with first HIV+ test at time of AIDS diagnosis to estimate rate of new infections among those who never test

# HIV Reporting Data

---

- ◆ Incorporate features of reporting system
  - *not* a random sample; people self-select to get tested at all
  - *only positive tests are recorded!!*
- ◆ Combine with detuned assay
- ◆ Note: number of new infections estimated from reporting data is *lower bound* on estimated total number of new infections

# Estimating HIV Incidence: Population Flow Approach

---

- ◆ Idea: develop a model of population flows that consistently defines:
  - What we want to estimate (the number of new HIV infections per unit time in the population that gets tested)
  - What we would observe in surveillance data (using the detuned assay on those who test HIV<sup>+</sup> via a sensitive antibody test)
  - Use the model to get what we want from what we see!

# Key Insights From Model

---

- ◆ Test frequency matters!! For given HIV incidence, you are more likely to observe those who test more rather than less frequently
- ◆ Obviously, you *never* observe those who never test – *must* look outside surveillance data to estimate incidence among such persons
- ◆ Known exposure time to HIV matters!! For given HIV incidence, those with longer times since last known HIV<sup>-</sup> test are more likely to be HIV<sup>+</sup> on current test, but less likely to be recently infected

# Key Assumptions

---

- ◆ By key population covariates (e.g. demographics, risk behaviors, location):
  - Constant HIV incidence in recent past (say last 2 years)
  - Constant individual average rate of testing  $\lambda$  (that is, on average test  $\lambda$  times per year);  $\lambda$  can vary by person
  - Can use detuned assay on HIV+ results
  - Only positive tests recorded
  - Intertest times independent of infection process
    - » Not as restrictive as it seems at first...more or less guaranteed to detect new infections among those who test and repeat due to recent risky exposure

# Basic Idea

---

- ◆ Use “prevalence = incidence x duration” to estimate number (prevalence) that are in detuned window period (i.e. recently infected)
- ◆ Modify to account for intertesting time
  - If intertest time is long relative to mean window  $E(W)$ , get “prev = inc x  $E(W)$ ”
  - If intertest time  $\tau$  is short relative to mean window, get “prev = inc x  $\tau$ ”
  - Account for testing rate
  - Observe number that detune (“prev”) and divide by duration to get an estimate of incidence!

# Details: What We Want

## (Start With Fixed Test Rate $\lambda$ )

---

- ◆ Want the number of new infections per unit time
- ◆ If population size is  $n$  and per capita incidence is  $r$ , we want

$$a = n \times r = \# \text{ new infections}$$

- ◆ If testing rate is  $\lambda$ , window period is  $W$ , and exposure time since last negative test is  $\tau$ , we see

$$s = \lambda \times n \times r \times E[\min(\tau, W)] = \# \text{ detuned tests}$$

from the detuned assay (remember stories for long versus short  $t$  compared to  $W$ )

$$a = s / \{\lambda \times E[\min(\tau, W)]\}$$

# Arrive At Result That Depends Only On Observed Data

---

- ◆ If you observe  $k$  recent infections in the surveillance data, and for each person you estimate the test rate ( $\lambda$ ) and the time since the last test ( $\tau$ ), then

$$\hat{a} = \sum_{i=1}^k \frac{1}{\lambda_i E[\min(\tau_i, W)]}$$

- ◆ Different ways to evaluate the above

# Have Developed 6 Different Versions of Above

---

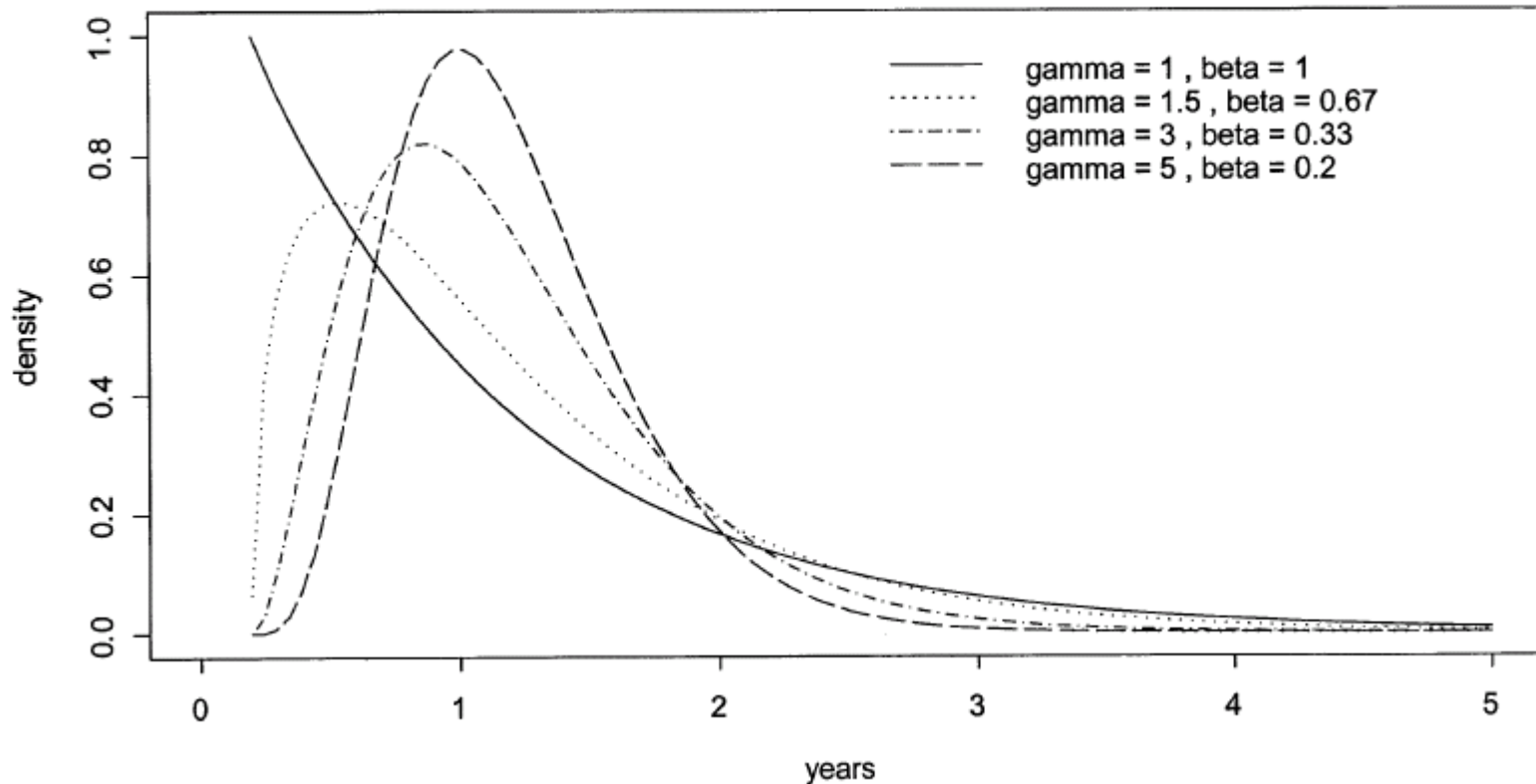
- ◆ For each, have formulas for estimated incidence and standard errors, enabling confidence interval estimation
- ◆ Evaluate via simulation of repeatedly testing population that produces 100 new infections per year on average
- ◆ Assume variety of intertesting time distributions

## Data on Testing Behavior (HITS, CDC, 2000)

	MSM (n=879)	IDU (n=823)	Hetero (n=926)
# of prior tests			
0	12%	12%	31%
1	11	13	19
2 – 3	27	36	30
4 or more	50	38	20
Median time since last test (months)	6	7	12

# Intertesting Time Distributions

Gamma densities for interest interval T  
minimum time, 0.2years;  $E[T] = 1.2$  years



# Simulation Results

Mean estimated incidence from 1000 simulations:  
 $E[I] = 100$ , gamma distribution for interest interval

Gamma params

$t_{\min}$	$\gamma$	$\beta$	E1	E2	E3	E4	E5	E6
0	1	1	100.1	100.3	102.8	102.1	107.1	92.3
0	2	0.5	98.8	100.4	107.7	102.6	104.3	97.4
0	5	0.2	100.3	101.1	106.8	103.7	104.5	102.4
0.1	1	0.9	100.3	100.9	106.4	102.7	106.7	96.6
0.1	2	0.45	98.8	101.5	106.3	103.1	104.2	99.5
0.1	5	0.18	99.2	101.1	105.4	102.5	103.1	101.6
0.2	1	0.8	100.7	101.1	109.3	103.8	106.6	100.3
0.2	2	0.4	99.5	100.1	107.2	103.2	104.1	101.1
0.2	5	0.16	99.5	103.2	105.6	102.9	103.3	102.3

# Simulation Results

Sample standard deviations of estimated incidence from 1000 simulations:  
 $E[I] = 100$ , gamma distribution for interest interval

Gamma params

$t_{\min}$	$\gamma$	$\beta$	E1	E2	E3	E4	E5	E6
0	1	1	17.7	61.5	15.9	19.7	16.5	14.2
0	2	0.5	16.1	40.0	15.3	17.8	15.2	14.2
0	5	0.2	15.7	34.6	15.2	16.2	14.9	14.6
0.1	1	0.9	17.1	39.0	15.9	18.7	16.0	14.5
0.1	2	0.45	15.6	33.5	14.9	16.4	14.6	14.0
0.1	5	0.18	14.8	34.0	14.6	15.2	14.3	14.1
0.2	1	0.8	16.7	32.3	14.9	18.2	14.5	13.7
0.2	2	0.4	16.4	31.5	15.6	17.1	15.2	14.7
0.2	5	0.16	15.2	36.3	14.8	15.4	14.5	14.3

# Simulation Results

95% confidence interval coverage (percent) from 1000 simulations:  
 $E[I] = 100$ , gamma distribution for interest interval

Gamma params

$t_{\min}$	$\gamma$	$\beta$	E1	E2	E3	E4	E5	E6
0	1	1	95	82	95	94	95	91
0	2	0.5	94	90	96	96	96	96
0	5	0.2	96	92	96	95	96	95
0.1	1	0.9	96	89	97	95	97	93
0.1	2	0.45	94	92	97	97	97	96
0.1	5	0.18	95	93	97	96	97	96
0.2	1	0.8	95	93	95	96	96	97
0.2	2	0.4	95	92	95	95	95	95
0.2	5	0.16	95	93	96	96	97	96

# Stuff That Can Go Wrong

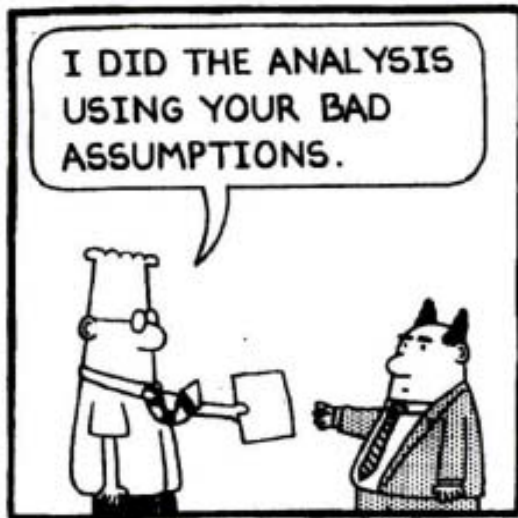
---

- ◆ New HIV diagnoses with no aliquot available for detuned assay
- ◆ Lab can't do detuned assay (currently test is not FDA approved, available only as IND, requires informed consent)
- ◆ Data needed that might not be available includes time since last test (or test frequency – linkage issue)

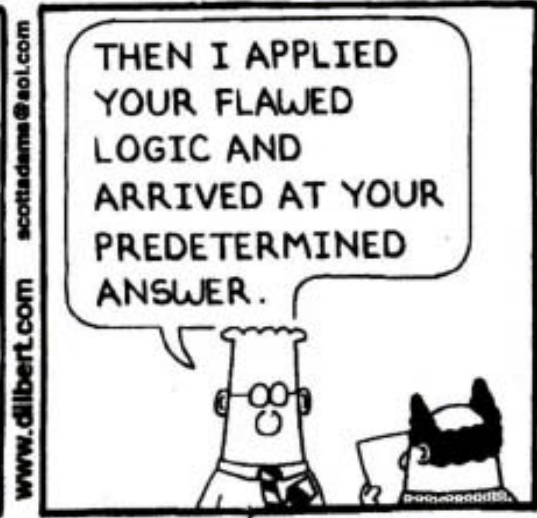
# Nonetheless...

---

- ◆ Simulations confirm models, and suggest that if the needed data become available, good incidence estimates can follow
- ◆ Of course, problem of estimating incidence outside reporting system remains



I DID THE ANALYSIS  
USING YOUR BAD  
ASSUMPTIONS.



THEN I APPLIED  
YOUR FLAWED  
LOGIC AND  
ARRIVED AT YOUR  
PREDETERMINED  
ANSWER.



SHALL I BEGIN  
DISILLUSIONING  
THE TEAM?

THIS NEEDS  
A PIE CHART.

www.dilbert.com scottadams@aol.com

3/30/00 © 2000 United Feature Syndicate, Inc.