Some thoughts on data collection in public health

Phillip Nieburg, M.D., M.P.H

CSIS Task Force on HIV/AIDS

April 18, 2006

Some data collection methods to guide public health practice

- **Surveillance: routine, active**
- **Surveys: population, cluster**
- **Disease registries**
- **Outbreak investigations**
- **Epidemiologic research**

Goals of Disease Surveillance

- **one of several diagnostic tools for epidemiology and for public health practice**
- **To "take the pulse" of disease spread**
  - **among individuals – eg., disease severity**
  - **in society – disease burden**
  - **for deciding on response policies**
    - **for prevention, care, mitigation, interrupt outbreak**
    - **for targeting, strength, duration of response**

Medical model analogy with public health data collection

- **Medical model: e.g., examining chest pain**
- **Many data sources**
  - **history (disease and family)**
  - **pulse, B.P., edema, cyanosis**
  - **stethoscope, EKG, chest x-ray, catheter**
  - **cardiac enzymes in blood**

- **Act on best guess for the moment**

HIV Prevalence

- Proportion of affected people in a population
- a snapshot
- Does not provide definitive information
  - Does not tell INCIDENCE (# of new cases)
  - Does not tell mortality (leaving population)
  - Does not tell disease burden (e.,g clinic beds)
  - Does not tell proportion of drug resistant HIV
  - Does not tell trend(s)

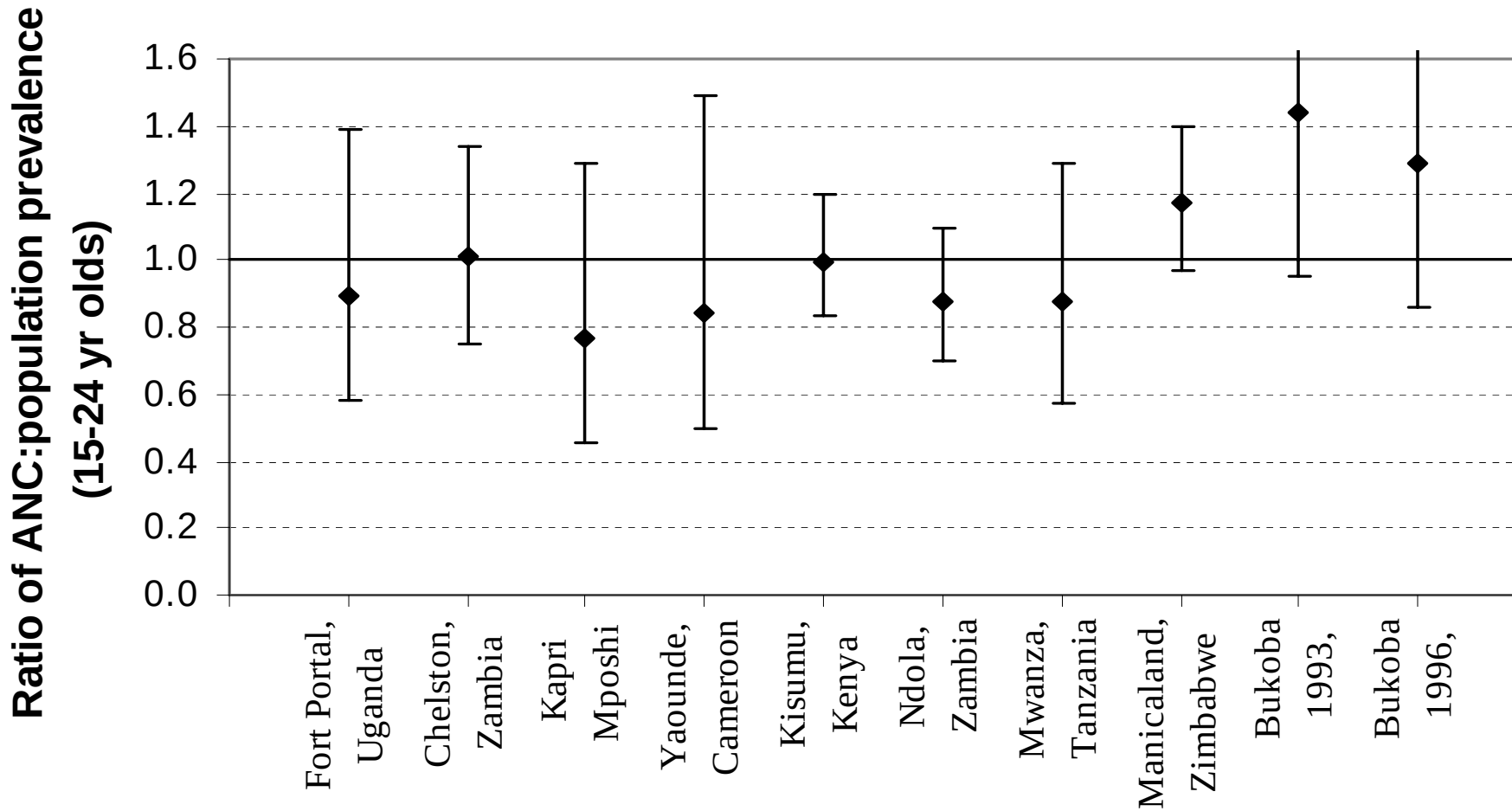
So why the mystique?

Potential Biases in ANC and Population-based Surveys

Issue	ANC	Pop. Survey	Other data suggest
absence	NA	Travel, illness, migration	Higher HIV missed
refusal	NA	High rates, esp higher SES	refusers more HIV; higher SES low HIV
urb.-rural	Urban	representative’	Rural w/less HIV
Women	All	‘representative	More HIV in older men, young women
pregnant	ALL	Many fewer	unprotected sex
pregnant	??	NA	HIV+ less fertile

Figure 3.4

Comparison of HIV prevalence among ANC attendees and the general population: Localized population-based studies



	Number tested	HIV-1 prevalence	Response rates		
			Household	Interviewed*	HIV-1 testing*
Zambia[20]					
All	3807	15·6	98·2%	92·9%	76·5%
Men	1734	12·9	..	88·7%	73·1%
Women	2073	17·8	..	96·4%	79·4%
Mali[21]					
All	6951	1·7	97·9%	90·0%	80·7%
Men	3069	1·3	..	83·8%	75·6%
Women	3882	2·0	..	94·9%	85·2%
South Africa[22]					
All	6080	15·6	71·1%	73·7%	62·1%†
Men	2585	12·8	..	..	58·4%†
Women	5361	17·7	..	..	65·1%†

*Numbers are those who were eligible to be interviewed or tested in households that agreed to participate. †Response rate for HIV-1 testing is for those aged 15 years and older.

HIV-1 prevalence and response rates by country and sex in respondents aged 15–49 years

from a 2003 W.H.O. document

Section 3.1.5. Estimating National Prevalence

“ANC prevalence data should not be directly extrapolated to the general population of the catchment areas, nor should regional or national median ANC prevalence be extrapolated to the general adult population.”

assumptions implicit in recent article and editorial

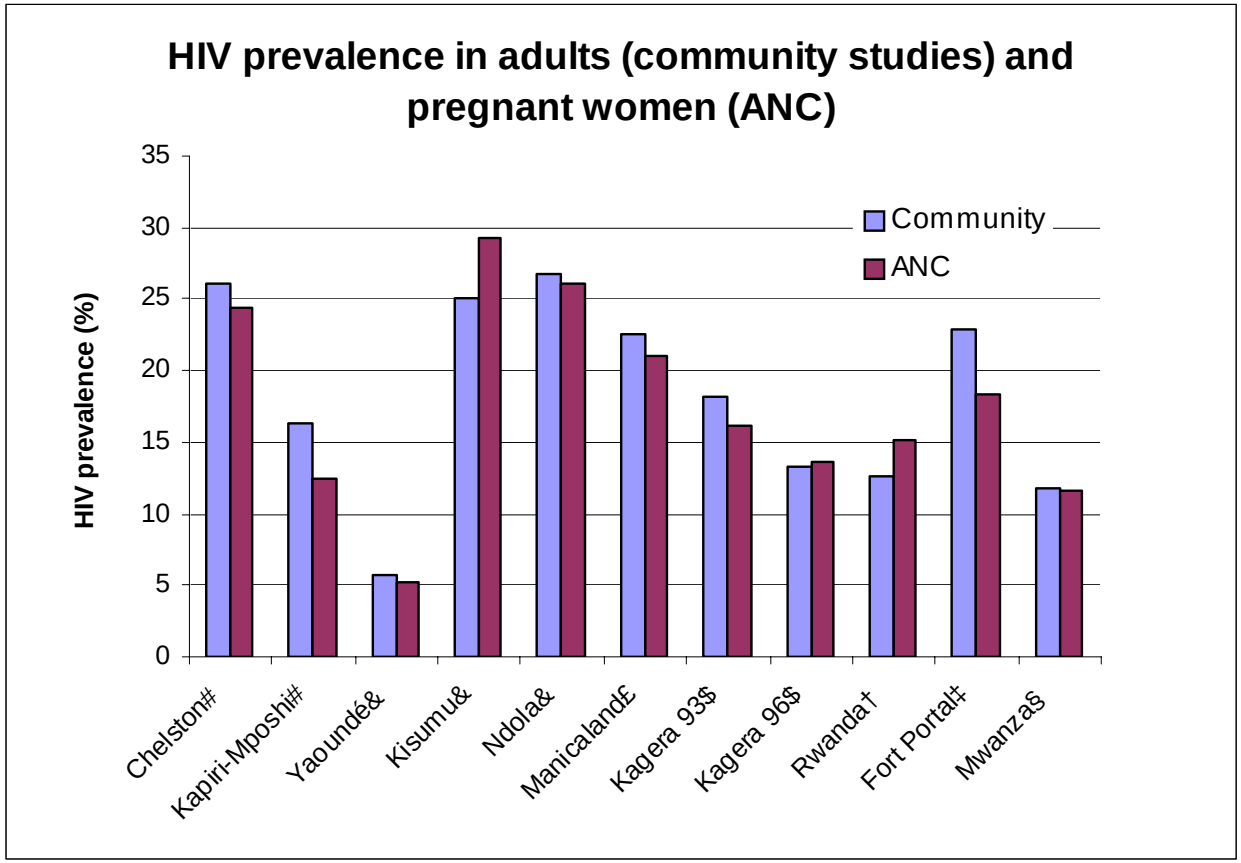
- ANC and population surveys have same purposes
  - and are “competing”
- there is “one right answer”
- ANC-survey discrepancy is new discovery
- Population based survey is “gold standard”

from a 2003 WHO document…

- **4. POPULATION-BASED SURVEYS WITH HIV TESTING**
- Population–based surveys differ in methodologies, sampling approaches, biological sample collection methods, HIV testing strategies, ways to deal with ethical issues and incentives for participation. These differences have to be taken into account when interpreting survey results.
- Non-response rates at the household level and at the individual level are major issues in interpreting the results. Refusal and absence are the main reasons and these are likely to have different associations with HIV prevalence.

HIV Survey-Surveillance Comparison, Zambia

Figure 3.2



comparisons on 15-39, adjusted; & marks comparisons on 15-50, age standardised
£ comparions on 15-44; $ marks comparions 15+, age standardised;
† comparisons on 15-44, non-adjsuted; ‡ comparison of crude data
§ crude data, community 15-54, antenatal clinic data 15+