Statistics 101 on GBV for practitioners: Understanding & Interpreting the Numbers
Making friends with quantitative findings

Kristin Dunkle PhD, Esnat Chirwa PhD & Shibe Mhlongo MS
What are the key questions we need to answer set policies and plan programmes?

- Prevalence: Who has the problem and how bad is it?
- Risk and Protective Factors: What causes the problem?
- Intervention Studies: How can we fix the problem?
Part 1: Prevalence and Incidence
How much GBV is Happening
Lifetime prevalence of GBV

- **Lifetime prevalence**: cumulative lifetime experience or perpetration to date
  - Most useful for:
    - Assessing need for services that are sensitive to survivors of violence
    - Studying medium to long term physical and mental health consequences of violence
    - Use in advocacy communications (tends to be highest number available)
    - Comparing geographic settings over the long term and/or historically

- Generally **terrible** for looking at impact of programmes or policies
- Subject to recall bias & therefore underestimates
Past year prevalence of GBV

- **Past year prevalence**: experience or perpetration of GBV over the past 12 months
  - Most useful for:
    - Getting a current snapshot of on-going problems with GBV
    - Comparing with most large-scale & standardized data sets
    - Assessing need for services that can serve newly-traumatized survivors of violence
    - Comparing with lifetime prevalence to see how violence persists or doesn’t
    - Looking at risk and protective factors for GBV
    - **Looking at impact of prevention programmes or policies**
  - Can be useful for advocacy communications (depends)
What about “incidence”?

• Classic disease-focused epidemiology:
  • **Incidence rate**, typically expressed in Number new cases per 100 or 1000 person years:
    \[
    \frac{\text{Number of new cases of disease during}}{\text{Total person-years of observation}}
    \]
    ex. Incidence of HIV among women aged 15-24 is 2.3 per 100 py
  • **Incidence proportion (or attack rate)**, typically expressed as a percent (%):
    \[
    \frac{\text{Number of new cases of disease}}{\text{Population at start of time interval}}
    \]
    ex. 10% of HIV patients become infected with TB each year
  • **Incidence proportion = period prevalence but going forward in time**
  • **Most commonly use metric in longitudinal GBV research & IE**
    i.e. What % of women experienced IPV on the 12 months since baseline?
    ex. At follow-up, 25% of women in the study reported experiencing of physical IPV in the past 12 months
Typical assessments in a GBV Impact Evaluation

• Baseline:
  • Lifetime experience or perpetration to date (lifetime prevalence)
  • Past year experience or perpetration (past year prevalence)

• 1 year post-baseline
  • Past year experience or perpetration (now serves as measure of incidence)

• 2 years post-baseline
  • Past year experience or perpetration (now serves as measure of incidence)

• Must always specify the time component in the follow-up measures
Confidence Intervals around Prevalence

• Quantitative studies are usually based on “samples,” not the full population.
• If the study were repeated using a different sub-sample of the population, the result might differ.
• Therefore, “Confidence Intervals” are reported with % (e.g. 95% CI):
  • If you did the exact same project 100 times, 95 times of those times would give a result inside the stated interval.
  • Ex. Prevalence of lifetime IPV among women in Zambia is 32.5 (95% CI: 31.9 – 33.1).
• A larger sample size usually gives a narrower confidence interval.
What do these look like?

- Point estimate: expressed as a %
- 95% confidence interval around the point estimate
- Is there a significant difference? Confidence intervals should not overlap

<table>
<thead>
<tr>
<th></th>
<th>Setting 1: South African Clinics % (95% CI)</th>
<th>Setting 2: USA National % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifetime prevalence of IPV</td>
<td>55.4% (52.3 – 58.7)</td>
<td>40.6% (37.7 – 43.2)</td>
</tr>
<tr>
<td>Past year prevalence of IPV</td>
<td>44.6% (38.4 – 53.6)</td>
<td>15.8% (10.6 – 20.2)</td>
</tr>
</tbody>
</table>
Assessing the quality of GBV data
Assessing a GBV estimate

Ask yourself 3 key questions about the research methods:

1. What was asked?
2. Who was asked?
3. How was it asked?
Assessing a GBV estimate: What was asked?

What was the GBV outcome of interest? How was it defined and operationalized?

Was a standard measure used?
If a non-standard measure, assess everything carefully

“Ask a stupid question, get a stupid answer”

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Some Recommended Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPV</td>
<td>WHO Violence Against Women Instrument (adapted for perpetration through extensive work in South Africa &amp; other settings), IMAGES, and the UN MCS of Men in Asia and the Pacific, recent DHS surveys</td>
</tr>
<tr>
<td>Sexual violence</td>
<td>IMAGES, UN MCS, What Works Sexual Experience Survey (SES) – N. America</td>
</tr>
</tbody>
</table>
Advantages of Benchmarking

• Questions on IPV drawn from WHO Violence Against Women Instrument (for women) and International Men and Gender Equality Survey (for men)
  • Used in over 16 countries by WHO
  • Used in 11 countries by IMAGES
  • Used in 6 countries UN Multi-country Study of Men in Asia and the Pacific
  • Model for current DHS module on domestic violence, used worldwide
• Allow international benchmarking of magnitude of program effectiveness
  • Used in Stepping Stones and IMAGE trials in South Africa
  • Used in SASA! in Uganda
  • All What Works trials
Assessing a GBV estimate: Who was asked? Sampling!

- Two key enquiries to make about sampling:
  - How was this sample drawn?
  - How does this sample compare to the population I want to talk about?

- Be cautious about over-generalisation!
  - (Authors do it, but readers and users of data do it more)
Generalizability: How do the people you are reading about compare to the people you want to talk about?

- Geographic area
- Age
- Socio economic status: Education, employment, housing, food security etc
- Relationship status (partnered, married, etc)
- Sexual orientation / gender identity
- Whatever else is relevant to your purpose
### Sampling Strategies and Generalizability

<table>
<thead>
<tr>
<th>Sampling strategy</th>
<th>Can be generalized to</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population-based household surveys</td>
<td>Valid for an entire geographic area</td>
</tr>
<tr>
<td></td>
<td>Examples: WHO Multi-Country Study, UN MCS, DHS, IMAGES</td>
</tr>
<tr>
<td>Other probability samples, using sampling frames</td>
<td>The population represented by the sampling frame</td>
</tr>
<tr>
<td></td>
<td>Examples: patients seeking care in a hospital system, students at a university, members of professional register</td>
</tr>
<tr>
<td>Respondent driven sampling</td>
<td>If done well, to members of a hidden population in the area covered</td>
</tr>
<tr>
<td></td>
<td>Examples: injection drug users, sex workers, men who have sex with men</td>
</tr>
<tr>
<td>Convenience</td>
<td>Not generalizable</td>
</tr>
</tbody>
</table>
Assessing a GBV estimate: How was it asked?

Fieldwork practices can have a huge impact on disclosure.
What is the single biggest problem in gathering any data on GBV?

Police reports
Reported to health or social services
Reported on a general survey
Reported on a specialized survey
What actually occurs

Getting people to tell you that it is happening

Detection Biases
• Underreporting bias
• Social desirability bias
Statistical measures for reading tables

- N: a count
- %: proportion of people to whom a count applies
- 95% confidence interval (95% CI):
- p or p-value: (roughly) how likely is it that difference/association seen in the data occurred by random chance and is not “real”.
  - Range $0 < p < 1$; smaller $p$ values mean more confidence
  - Usual thresholds: $p < .05$; $p < .01$; $p<.001$ (etc)
  - Careful when doing many simultaneous comparisons
- Mean: average value
- Median: Most common value
Questions ?
Part 2: Understanding and Using Data on Risk Factors & Protective Factors for GBV
Where do we get information on risk & protective factors for GBV?

**Observational Studies**

<table>
<thead>
<tr>
<th>Cross sectional surveys</th>
<th>Longitudinal/Prospective surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Can estimate prevalence (may or may not be representative)</td>
<td>• Follow the same group of people over time</td>
</tr>
<tr>
<td>• Can suggest risk and protective factors</td>
<td>• Can establish incidence</td>
</tr>
<tr>
<td>• <strong>Cannot</strong> establish causation</td>
<td>• Can help establish causation (know time sequence of events)</td>
</tr>
<tr>
<td>• Very commonly used for GBV – now have multi-national datasets to look across settings</td>
<td>• UNCOMMON in GBV research – it’s unethical to refrain from helping!</td>
</tr>
</tbody>
</table>

**However! You CAN use data from intervention studies to answer observational questions**
Where do we get information on risk & protective factors for GBV?

- Primary or secondary analysis of data from impact evaluations
- Follow the same group of people over time
- Can establish incidence
- Can help establish causation (know time sequence of events)
- Becoming MORE common in GBV research – lots of prevention trials!
Some Common Terminology

**Outcome variable (Y)**
- The thing you hope to change or most want to explain (for most of us, GBV)
- aka dependent variable
- aka criterion variable

**Independent variables (X₁, X₂, X₃, ... Xₘ etc)**
- The intervention in experimental research
- aka predictor or exposure variables in epidemiological studies
- aka correlates in cross-sectional studies
- aka risk factors, protective factors, drivers
3rd Variables (usually in models only)

- **Mediators** (mediating variables, effect modifiers)
  - A variable that explains (or mediates) the observed relationship between two other variables
  - Changes the of the relationship between X and Y (usually by getting in the middle)
  - **Example**: The association between experience of sexual violence and increased alcohol use is mediated through and increase in depression & PTSD

- **Moderators** (moderating variable, buffer, confounder)
  - Identifies conditions under which the relationship between two variables might be stronger or weaker
  - Also: covariates, extraneous variable, lurking variable
  - **Example**: Alcohol increases risk of non-partner sexual violence for young women at university more than older married women
### Example: Comparing age and education

<table>
<thead>
<tr>
<th></th>
<th>Women (N=1400)</th>
<th></th>
<th>Men (N=1400)</th>
<th></th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% CI</td>
<td>Mean</td>
<td>95% CI</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Age</td>
<td>33.0</td>
<td>(32.4-33.6)</td>
<td>34.8</td>
<td>(34.2-35.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>No school</td>
<td>240</td>
<td>17.2</td>
<td>241</td>
<td>17.2</td>
<td>0.74</td>
</tr>
<tr>
<td>Primary</td>
<td>911</td>
<td>65.2</td>
<td>913</td>
<td>65.2</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>158</td>
<td>11.3</td>
<td>163</td>
<td>11.6</td>
<td></td>
</tr>
<tr>
<td>&gt; Secondary</td>
<td>60</td>
<td>4.3</td>
<td>62</td>
<td>4.4</td>
<td></td>
</tr>
</tbody>
</table>
More on p-values and confidence intervals

- P-values and Confidence intervals are based on measures of the “standard error”
- The standard error tells us how precise our measurements are
  - Note that it tells us nothing about accuracy!
- There are different ways to calculate standard error if you have
  - Simple random sample
  - Stratified sample
  - Cluster sample
- There is no single best approach
Suppose we want to know how STRONGLY a factor is associated with IPV.

Most common measures are Odds Ratios and Risk Ratios or Relative Risk.

If you want to understand them, take an intro epi class.

If you just want to USE them ....
## Dr. Dunkle’s quick and dirty guide to odds ratios risk ratios, part 1

<table>
<thead>
<tr>
<th>Odds Ratios</th>
<th>Risk Ratios / Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Come from cross sectional data</td>
<td>(Usually) come from longitudinal data</td>
</tr>
<tr>
<td>Speak to correlations</td>
<td>Aspire to be about causality</td>
</tr>
<tr>
<td>Often OVERESTIMATE associations</td>
<td>Are usually pretty accurate</td>
</tr>
<tr>
<td>Odds of 1:1 gives a ratio of 1.00 = no effect</td>
<td>If risk in exposed = risk in unexposed, then Risk (exposed)/Risk (unexposed) = 1.00 = no effect</td>
</tr>
<tr>
<td>OR &gt; 1 = positive associations = risk factor</td>
<td>RR &gt; 1 = positive associations = risk factor</td>
</tr>
<tr>
<td>OR &lt; 1 = neg association = protective factor</td>
<td>RR &lt; 1 = neg association = protective factor</td>
</tr>
<tr>
<td>Can be crude (2 variable) or adjusted (in a model with lots of variables)</td>
<td>Can be crude (2 variable) or adjusted (in a model with lots of variables)</td>
</tr>
<tr>
<td>Use the words “increased odds / decreased odds” to discuss. Do NOT say “risk.” Be smug.</td>
<td>Go ahead and talk about “risk”. It’s right there in the name.</td>
</tr>
</tbody>
</table>
Build Models

What is a model?
- A set of calculations that account for relationships among many variables simultaneously
- A set of statistical relationships that (hopefully!) represent some aspect of reality

Why model?
- Everything in your data is interdependent; models help handle this complexity
- Eliminate “confounding”
  - *Example:* Coffee drinkers get more lung cancer. Does coffee cause lung cancer? NO. But coffee drinkers are more likely to smoke cigarettes.
- Learn which risk & protective factors have biggest impact on outcome
  - *Does education or availability or services matter MORE?*
Commons Models in IPV Research

Multiple or Multivariable Regression: Multiple variables pointing together at a single outcome variable

**Logistic regression: dichotomous outcome**
- IPV No or Yes
- Expressed as Odds Ratios (sometimes Risk Ratios)

**Multinomial regression: categories**
- IPV: None, Physical Only, Sexual Only, Both
- Each other category compared to “none” in turn
- Often expressed as Risk Ratios

**Poisson (or log-linear) for counts**
- IPV: How many times did it happen?
- Often expressed as coefficients
Commons Models in IPV Research

Hierarchical Linear Models (see also Generalized Estimating Equations)

- **Super Great** for cluster sampled data
- Equally good for longitudinal or cross-sectional
- “Hierarchical” means looking at the data at multiple different levels at the same time

**Cluster Level**
- Intervention or Control
- Aspects of the area
  - # Hospitals
  - # Police Stations
  - Distance from capital
  - Access to improved water

**Person Level**
- Things that don’t change:
  - Male or Female
- Things that happened before the study:
  - Education Completed
  - Child abuse

**Time-varying**
- Sexual partners in past year
- Marital Conflict past year
- Alcohol use in past year
- Hunger in past month
- Depression in past week

IPV at most recent measure
Commons Models in IPV Research

Structural Equation Modelling (also path analysis)

- Looks at multiple pathways between a collection of variables
- Can include “latent” variables – multiple scale items pointing at a single construct
- Output is a coefficient for each line and a test for significance; non-sig lines are usually dropped
- Good for cross sectional or longitudinal data
- MUST be solidly grounded in a theory
- Requires specialized methods to do this on cluster sampled data, but can be done
Useful points and tips about models

• Building from a THEORY almost always helps

• In general, more “parsimonious” (smaller and simpler) models are preferred, even if you can get the same answer in a more fancy way.

• If you get the same answer by doing the model more than one way, your answer is more likely to be real (consistency, coherence).

**Top Tip:** Always, always, always check the N (number of observations) included. If more than a handful are missing, this is a problem.

---

**Most often attributed to McCullagh and Nelder (1983, 1989)**

1. All models are wrong, but some are more useful than others.

2. Do not fall in love with one model to the exclusion of others.

3. Thoroughly check the fit of a model.
Part 3: Understanding and Using data from RCTs and other Impact Evaluations
What is impact evaluation?

- There are many types of “evaluation”
  - Needs assessment: What do our stakeholders needs or want? What is missing?
  - Formative or developmental: How might we fill the gaps?
  - Process or operational: Are we doing what we said we would? How well is it going?
  - **Impact evaluation:**
    - Did things change?
    - Are they better? Are they worse?
    - *Would they be the same even if we did nothing?*
  - Academic & research types often talk about “trials”
Impact Evaluations are a special case for risk factor analysis

- Key protective factor of interest is assigned
- Generally looking for Risk Ratios because longitudinal
  - You are trying to DISPROVE $RR = 1.00$
  - Protective effect is $RR < 1.00$ for the intervention
  - Confidence interval must NOT include 1.00
  - P-value if presented should be significant
  - RR should be adjusted for any factors that differed at baseline
  - Can interpret 1-RR as % reduction in outcome
    - i.e. $RR= 0.55$ (95% CI 0.45 – 0.65) = 45% reduction in GBV
Key elements of impact evaluation I

What do change we want?

- Define a set of key outcome variables / indicators based on objectives

Where we are we starting?

- Assess or measure key variables / indicators BEFORE starting (Baseline / Pretest)

Did things change? Are they better? Are they worse?

- Must look at multiple time points
Where did we get to? What’s different?

- Assess or measure key indicators AFTER program (Endline / Post-test)
- If interested in the process / progress of change, can measure 1 or more times DURING program (Midline)
- If interested in sustained impact, can measure again 1 or more times AFTER the end of the program
Would things be the same even if we did nothing?

- If we want to say our program caused change, we need to compare our outcomes to people who have all the same experiences EXCEPT our program.
- Requires a counterfactual or “control” population:
  - The world changes all the time.
  - People we are targeting do things for themselves.
  - People who are not us offer services.
  - Big events happen that impact everyone.
Type of study designs

**NON-EXPERIMENTAL**
- Measure outcomes before and after program for participants only
- No comparison group

**QUASI-EXPERIMENTAL**
- Measure outcomes for program participants and non-participants without random assignment
- "Control" for bias
- Comparison group

**EXPERIMENTAL/RCT**
- Randomized Control Trial (RCT)
- Randomize participants to treatment or control group
- Measure outcomes for both groups
- Explicit comparison group

*Increasing rigor*
Randomized Controlled Trials (RCTs)

- Randomized Controlled Trials
  - “Gold standard” of evidence
  - Individuals or groups of people who are equally eligible and split by chance into intervention and control groups
  - Allocating by chance eliminates volunteer and selection bias
  - Eliminates history bias (i.e. everyone experiences the same current events)
  - Allows change to be attributed to the programme

- Most persuasive to policy makers and donors
- Best (potential!) tool to advocate for program effectiveness
RCTs, Individual Randomization

- Participant enrols and then is randomly allocated to intervention or control
- Statistically simplest, requires smallest sample sizes
  - Often less expensive
- Must be able to logistically handle individual allocation, e.g.
  - Randomly allocate clinic patients to new counselling strategy or standard care
  - Randomly allocate students to different versions of a programme
- Works best with interventions that are delivered to individual people & interventions where little spillover/contamination to controls is expected
RCTs, Cluster Randomization

- Most common in GBV research
- Used when intervention is delivered at group level or expected to have group effects
  - By geography: village, province
  - By institution: clinic, school
  - Other logistical reasons
- Groups randomly assigned to intervention or control
- Statistically more complex, must adjust sample sizes and analysis to account for the clustering
Key Variations in Impact Evaluation

- **Non-Experimental**
  - No control groups or counterfactual
  - Can select people by any strategy
  - Measures (hopefully) at multiple points in time
  - Is easiest and cheapest
  - Cannot say that change was due to program

- **Quasi-experimental**
  - Uses a control or counterfactual group, but not fully random
  - Usually easier and cheaper
  - Eliminates history bias; vulnerable to selection & volunteer bias
Evaluation without impact

- Endline or post program assessment only
  - Good for assessing participant impressions of and reactions to a programme
  - Good for demonstrating feasibility or acceptability of an intervention strategy
  - Can be very useful in looking at process and operational variables

- Cannot assess impact
Qualitative Evaluation

• Looks in depth at smaller numbers or people or groups
• Able to gather deep and detailed information about process and motivations for change
• Useful what quantitative methods are infeasible (too expensive, too complicate)
• Useful when process of change is not yet understood
• Best for answering questions about “how” and “why”
How to learn more...

Suggested resources (NOT official endorsements):

• David Kleinbaum (legendary prof!) free, low-bandwidth, online class in epidemiology: http://activepi.herokuapp.com/

• Intro Epidemiology online by UNC faculty; free if you do not need a certificate https://www.coursera.org/learn/epidemiology:
  • Coursera & other MOOCs also have social science stats courses
  • Adding courses regularly in languages other than English

• Mess around with data and make as many mistakes as you can!

• Prof. Google is your best friend
  • Put is the name of your software and what you want to do: someone has written code!
  • Lot of useful tutorials can be discovered this way