THE INTER-SECRETARIAT
WORKING GROUP ON
HOUSEHOLD SURVEYS

Chapter 8: Weighting

UN Handbook of Surveys on Households and Individuals

Diego Zardetto & Kevin McGee (World Bank)

ISWGHS Handbook Webinar Series · 24 September 2026

Outline

1. What are survey weights and why they matter (§8.1)
2. Building the weights, step by step (§8.2)
3. Beyond the standard process: panel, pooled & non-probability samples (§8.3–8.4)
4. Country case study & key takeaways (§8.5)

ISWGHS

PART 1 OF 4

8.1 Foundations

Why survey weights matter, and the core three-step process that produces them

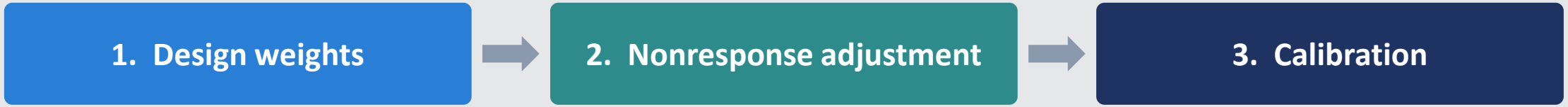
What are survey weights and why they matter

- Survey weights are the **bridge** between the sample and the target population
- **Weighted estimators** multiply each unit's data by its survey weight to produce population estimates
- Most household surveys use complex, **multistage stratified designs** with **unequal inclusion probabilities**
- **Unweighted estimators** (e.g. the plain sample mean) are generally **biased** under these designs
- Weights are a critical, integral component of survey data quality — **not an afterthought**

ISWGHS

Objectives and the weighting process

- Goal: build **estimators** that are **unbiased** (or nearly so) and **as precise as possible**
- A hierarchy of requirements:
  1. **Achieve unbiasedness** under *ideal conditions* (design weights achieve this)
  2. **Minimize bias** from *non-sampling errors* (nonresponse, frame undercoverage)
  3. **Leverage auxiliary information** to mitigate bias and improve precision
- All three objectives are rarely achievable in a single step → a sequential, multi-step process with **three core steps**:



  - Each step should improve the weights it receives **without undoing the gains** of the step before it

ISWGHS

Guiding principles for every adjustment

- Every adjustment step should be:

Principled
With a clear statistical goal

Controlled
Using established, rigorous methods

Minimal
The smallest change needed (to achieve the statistical goal)

- Heuristic or subjective modifications MUST be avoided — they would **damage inference**
- Weights must be **"universal"**: one set of weights per *analysis unit*, usable for all estimates regardless of the interest variables
  - Each time an additional selection/subsampling occurs, a new analysis unit is generated, which requires a specific set of weights
- When units are linked (e.g. households, individuals, plots, enterprises), **"integrated weights"** keep estimates logically consistent across levels

ISWGHS

PART 2 OF 4

8.2 Building the weights, step by step

A practical walk-through of the core steps, with a focus on multistage household-survey designs

Step 1 — Design weights

- **Design weight** = reciprocal of the unit's sample inclusion probability (the “base” or “initial” weight)
- Builds directly on the sampling design:
  - E.g., SRS, stratified, PPS
  - For a typical household survey: **multistage stratified cluster design**
- Ensures **unbiasedness under ideal conditions**
- Therefore, it serves as the **mandatory starting point** for every later adjustment

ISWGHS

Step 1.5 — Adjusting for unknown eligibility

- Some sampled units may turn out, only at the point of interview, to be ineligible
  - → More concerningly, for some units – particularly nonrespondents – **eligibility may be unknown**
- In those cases, an **unknown eligibility adjustment** is necessary before nonresponse and calibration
  - Basically, the weights of the units whose eligibility status is known are inflated to account for units with unknown eligibility
- This early adjustment step **prevents bias** while allowing users to restrict the analysis to units that are **known to be eligible**

ISWGHS

Step 2 — Nonresponse adjustment

- Addresses **unit nonresponse** through weighting (item nonresponse is handled separately, via imputation)
- Two common approaches:
  1. **Response homogeneity groups (RHGs)**
  2. **Logistic response-propensity model**
- Adjustment factor = reciprocal of the estimated response propensity — **always ≥ 1**
- Needs special treatment in **multistage designs**, where nonresponse can occur at more than one sampling stage
- **Field substitution** during fieldwork is **not a substitute** for this step — it does not guarantee protection against nonresponse bias

ISWGHS

Step 3 — Calibration

- **Calibration weights** make survey estimates exactly match known population totals (“auxiliary variables”) from external sources
- The adjustment is **minimal by design** — it doesn't undo the nonresponse-bias protection already built into the weights
- **No closed-form solution** in general — computed via iterative numerical optimization
- Two familiar special cases:
  1. **The post-stratified estimator**
  2. **The raking estimator**

ISWGHS

Calibration: choosing auxiliary variables

- Good auxiliary variables do one or both of:
  - **Improve precision**
  - **Mitigate nonresponse or undercoverage bias**
- Population totals can come from **censuses, administrative registers**, or other high-quality external sources
- **"Integrated household–individual weights"** keep estimates logically consistent across linked units (e.g. household size × households = individuals)
  - Achieved by setting calibration constraints **simultaneously** at the household and individual level

ISWGHS

Trimming extreme weights

- Nonresponse and calibration adjustments **increase weight variability** — sometimes producing very large (or even negative) weights
- Extreme weights can result in **unstable estimates**, especially for skewed variables like income or consumption
  - Trimming extreme weights reduces weight variability and counteract unstable estimates — but at a cost (**bias**)!
- Trimming is **heuristic**, not grounded in mature theory, and **risks introducing bias**
- **Apply sparingly** — only after:
  1. Assessing whether it is in fact necessary and,
  2. Carefully evaluating the trade-off between precision gain and potential bias injection

ISWGHS

Weight-calculation diagnostics

- Diagnostics help ensure that the final weights are **valid, stable, and trustworthy**.
- Two classes of checks:

Definitive checks

- Confirm the weights meet theoretical requirements
- Any failed check identifies a mistake

Quality-monitoring diagnostics

- Compare weighted estimates before and after adjustments and against external benchmarks
- Flag extreme or influential weights

- Diagnostics can guide decisions on **trimming** and other corrective measures
- An essential quality-control measure to be conducted **at each adjustment step** as well as **before disseminating** the weights

ISWGHS

Disseminating and documenting weights

- Weights must be **“universal”** — a single set per analysis unit, usable for the full range of survey estimates
  - Document how the weights of **linked analysis units** (households, individuals, enterprises, plots) are related
  - If they are **integrated**, explicitly mention it
  - **Avoid** cosmetic adjustments to the weights (e.g., *rounding*, *normalizing*)
- Document every step, including:
  - Definitions and goals
  - Adjustment methods and any associated auxiliary variables
  - Population benchmarks used for calibration
  - Ideally, also the diagnostics performed
- **Transparent metadata** builds **credibility, reproducibility**, and **correct use** by data analysts

ISWGHS

PART 3 OF 4

8.3–8.4 Beyond the standard process

Weighting for panel and pooled surveys, and extending the toolkit to non-probability samples

Panel surveys and pooled samples

Panel surveys

- **Cross-sectional weights**
  - Rebuilt each wave - From wave 2 they start from the prior wave's final weights, not new design weights - Adjustments are critical to maintain representativeness over time (e.g. to counteract *attrition*)
- **Weight-share method**
  - Adjusts for changes in household composition between waves - *Reduces* the weight of households gaining new members, as they might have been selected in *multiple* ways
- **Longitudinal weights**
  - Needed jointly analyze linked data from different waves (*longitudinal population*) - They enable estimation of *gross-change* across waves (e.g. state transitions)

Pooled samples

- **Combine independent samples**
  - Often to increase estimate *precision* or better analyze *rare subpopulations*
- **Practical applications**
  - Combine quarterly LFS rounds to compute yearly estimates - Expand a panel survey adding "refreshment" or "top-up" samples - Conduct multiple-frame surveys using overlapping frames
- **Pooling adjustment**
  - Units in the combined sample might have been selected in *multiple* ways - To compensate for this, the pooling adjustment *reduces* their original weights - Use of the original weights would lead to *bias* - Optimal adjustment factors depend on the *effective size* (n / design effect) of the original samples

ISWGHS

Extending the toolkit to non-probability samples

- Non-probability samples lack the controlled randomness needed to calculate **design weights** — and therefore are necessarily **exposed to bias**
- **Data-integration methods** let them "borrow representativeness" from a probability sample or census, via shared "bridge" variables
- Two classes of approaches:

Modelling the weights

- Propensity modelling
- Pseudo-calibration

Imputing the interest variable

- Sample matching
- Mass imputation

- Achieving unbiased estimation hinges on **strong assumptions** about the bridge variables, which are **very difficult to test**
- **Probability sampling remains the dominant paradigm** in official statistics

ISWGHS

Country case study: Ethiopia ESPS-5 (2021/2022)

- Describes the **full cross-sectional weighting** process in a large-scale panel household survey:



- Shows the rationale, working mechanism, and practical impact of each adjustment step on the **weight distribution**
- A worked example illustrating the **challenges** practitioners often face in low- and middle-income countries, as well as **best practices** to address them

ISWGHS

PART 4 OF 4

Key Takeaways

Key takeaways

- **Design weights → nonresponse adjustment → calibration**: a principled, controlled, minimal sequence
- The same logic extends, with care, to **panel data** and **pooled samples**, as well as to **non-probability sources** (though with strong limitations)
- **Diagnostics and documentation** are **not optional extras** — they are part of producing trustworthy weights
- Weight calculation should NOT to be undertaken as an algorithmic process
  - Each survey has its **own peculiarities** that should be carefully considered and addressed **using validated methods**
- Chapter 8 is meant as both a **step-by-step guide** and a **reference for best practice**
  - Highlights core **principles** and provides **methods** that can serve as **ingredients** in a weight calculation process — **not a definitive one-size-fits-all process**

ISWGHS

Thank you

unstats.un.org/iswghs