THE INTER-SECRETARIAT
WORKING GROUP ON
HOUSEHOLD SURVEYS

Positioning Household Surveys for the Next Decade

Gero Carletto
Co-chair, Inter-Secretariat Working Group
on Household Surveys

unstats.un.org/iswghs

The ISWGHS: a primer

- ❑ Established in 2015 under the aegis of the UNSC
- ❑ Objectives:
  - ❑ Improve coordination of household surveys
  - ❑ Advance cross-cutting survey methodology
  - ❑ Enhance communication and advocacy
- ❑ Governance
  - Membership: 11 international agencies + 8 (rotating) member states
  - Secretariat: UN Statistics Division
  - Current co-chairs: WB and UNW
- ❑ Work through time-bound Task Forces, led by and with contribution from members and non-member experts.

ISWGHS

A Thing of the Past?
Household Surveys in the Era of Big Data

Gero Carletto
Development Data Group
The World Bank
gcarletto@worldbank.org

President's Invited Lecture, International Association of Survey Statisticians (IASS)
World Statistics Congress, Kuala Lumpur, Aug 18-23, 2019

WORLD BANK GROUP
Development Data Group

Are household surveys a relic of the past?

WORLD BANK GROUP
Development Data Group

Big Data, Bigger Hype?

Big Data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it.

Dan Ariely
Professor of Psychology & Behavioral Economics

WORLD BANK GROUP
Development Data Group

New Data sources [and affordable technologies] as an opportunity

- Burgeoning literature on data integration
- More sophisticated tools and methods exist, and growing
- [Fastly evolving, more affordable technologies]
- New role for household surveys!

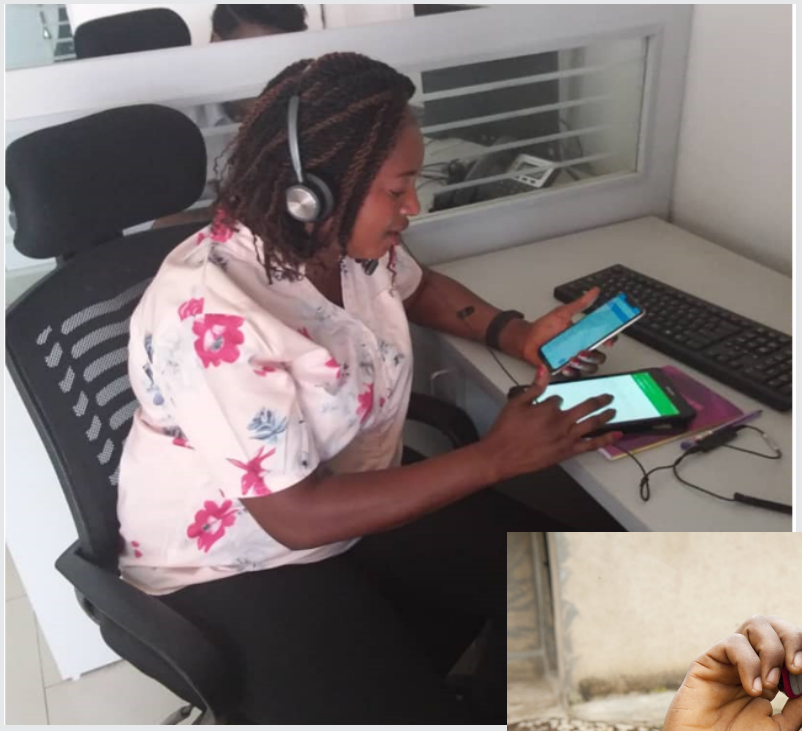
Adding value to surveys and address (mostly fair and constructive) criticisms …

- Make it more relevant!
  - Tool for **validation/calibration** of alternative data sources
    - Successful integration requires some repurposing/"add on"
      - At design stage
      - Interdisciplinary fora, "common language"
- Make it more accurate!
  - From self-reported to **objective** measurement
    - E.g. land, soil, water quality, varietal identification, food consumption, CAPI, …
  - Use Big Data (eg paradata, gridded PopMaps) to improve accuracy
- Make it more cost-effective!
  - [Through greater **integration**]
  - Objective measurement cum **imputation**
  - But also, more attention to exploiting and quantifying **benefits**!

In less than a year from KL, we went …

… from this …

… to this!



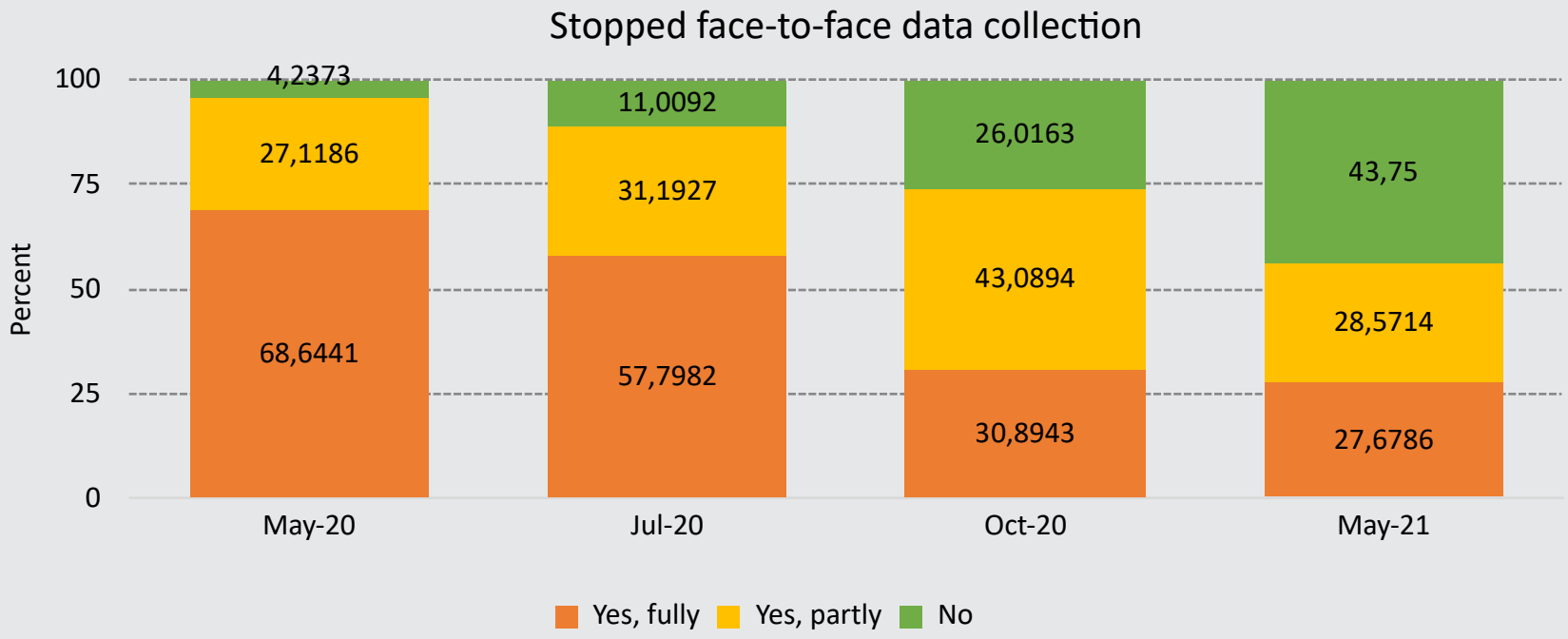
ISWGHS

From this …

… to this!

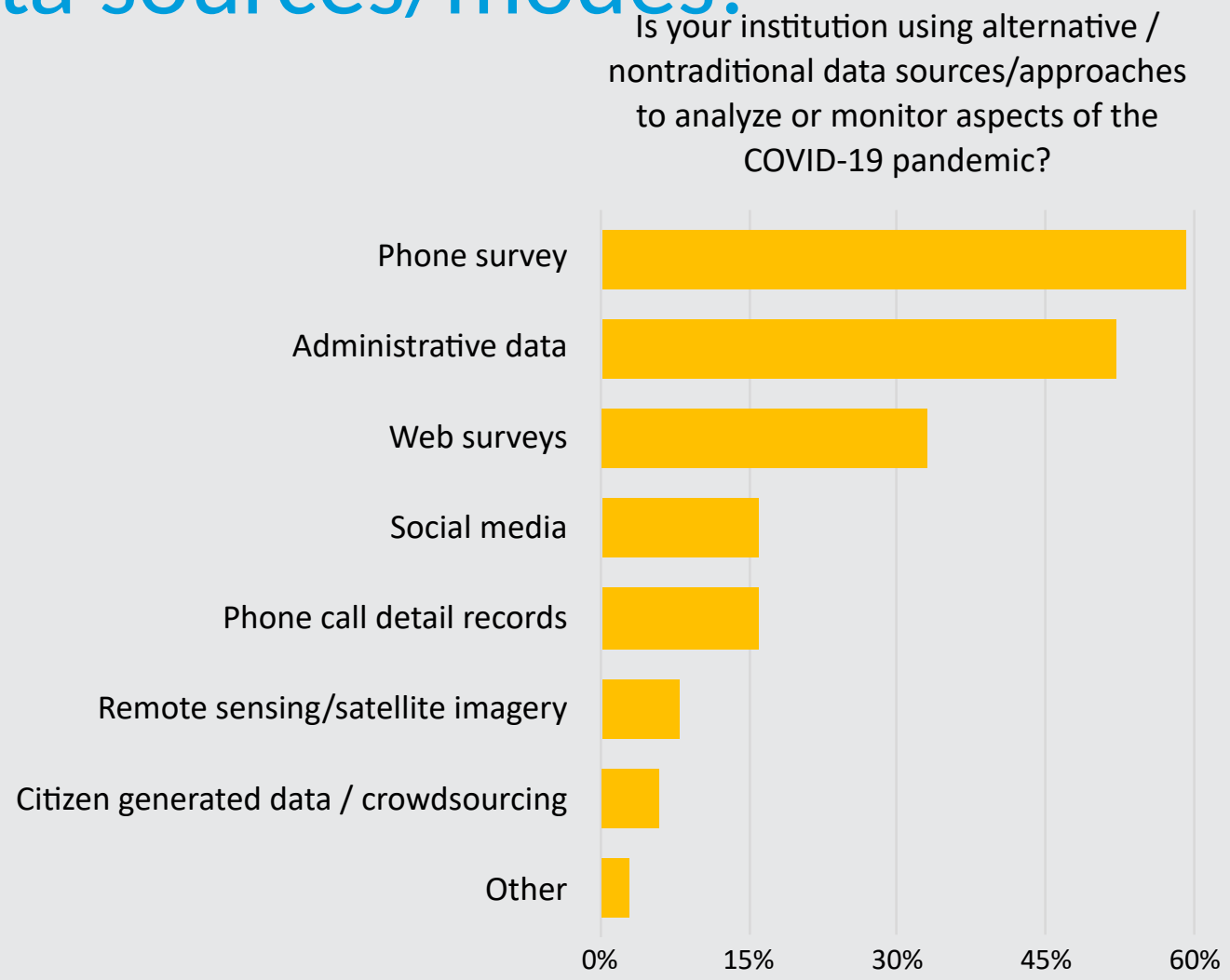
ISWGHS

As of May 2020, virtually all countries had fully/partly stopped F2F surveys …



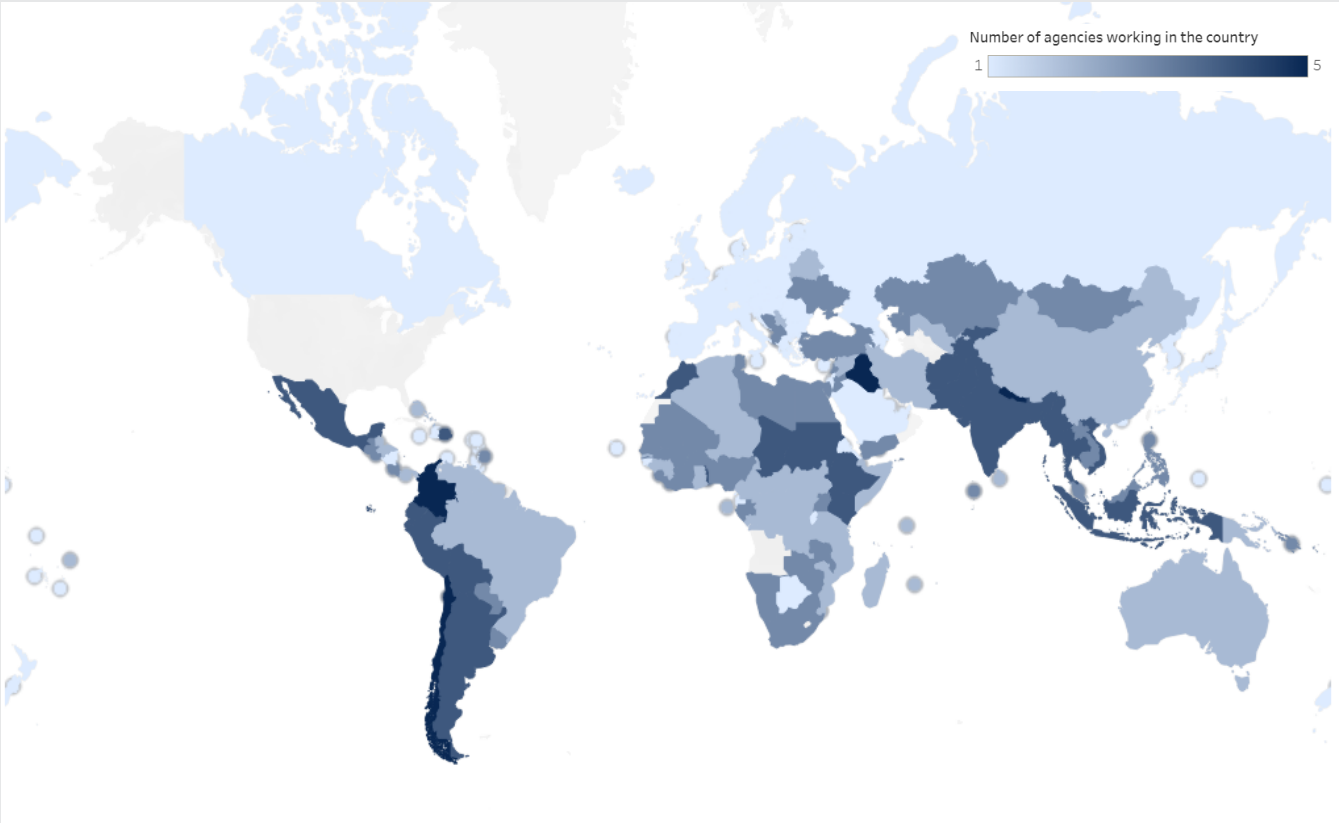
ISWGHS

… but many NSOs responded rapidly by adopting new data sources/modes!



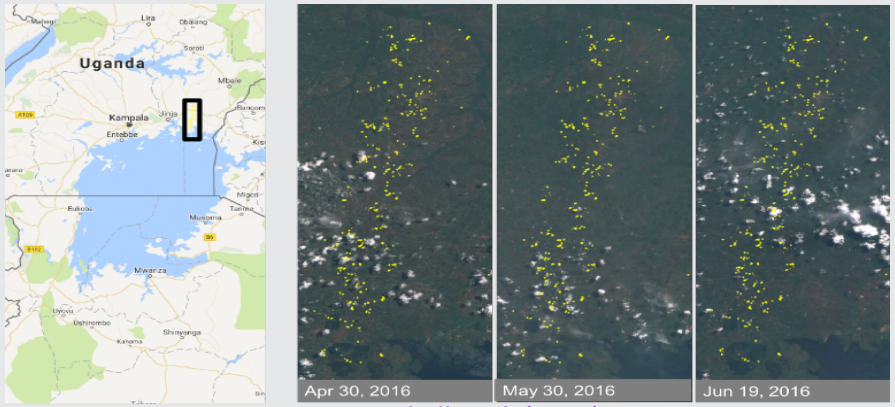
ISWGHS

Virtually all countries are now conducting phone surveys to monitor impact of pandemic and beyond, most with support from ISWGHS members …

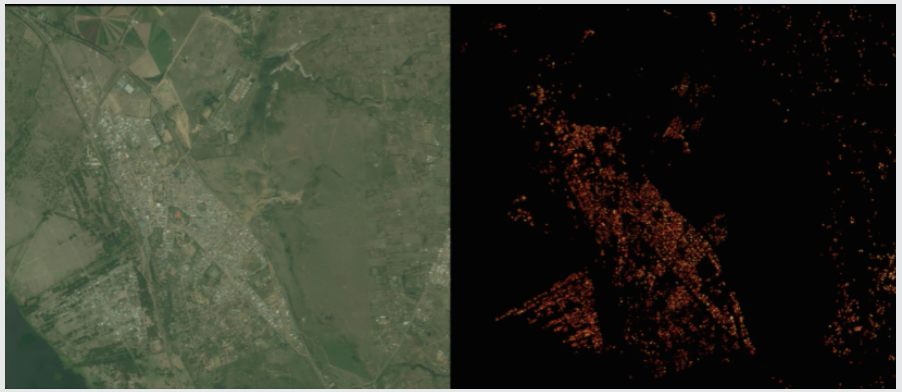


ISWGHS

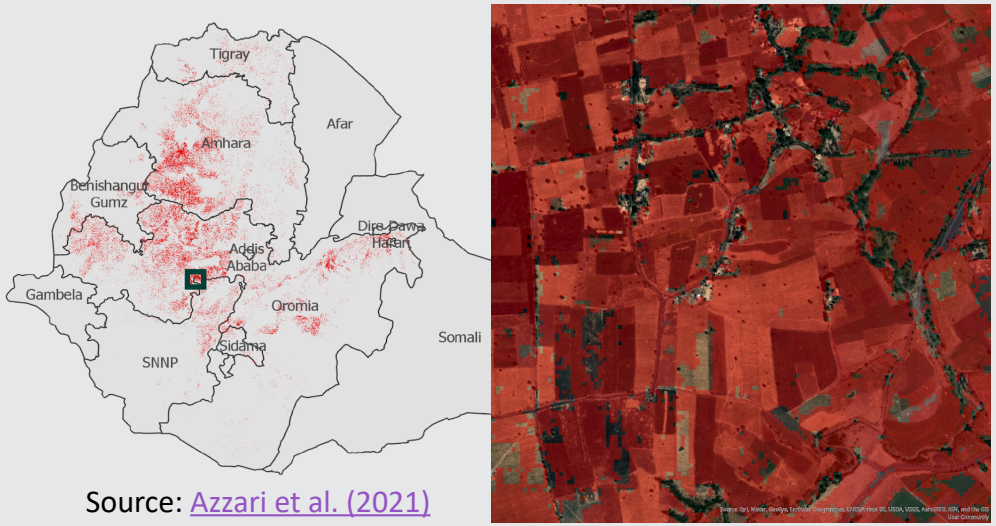
… and many are integrating EO and survey data and leveraging ML to improve spatial disaggregation and timeliness of development data …



Source: Lobell et al. (2020)



Source: Tiecke et al. (2017)



Source: Azzari et al. (2021)

ISWGHS

… meanwhile, affordable sensors and other digital instruments are increasingly used in survey operations.

Source: Sinha et al. (2020)

Source: Carletto et al. (2017)

Source: Friedman et al. (2021)

ISWGHS

However …

- Lack of sound survey infrastructure: only 43% of 180 countries implementing phone surveys used a recent, updated sampling frame
- Institutional barriers: NSS not set up to foster interoperability
- Lack of financial and technical capacity: main concern voiced by (L/MICs) NSOs in UNSD/ WB survey
- More often, data integration and correction for bias done ex-post with data which are unfit-for-purpose
- EO/modelled estimates of crop production and other remote-sensed applications reveal huge differences with ground-based, “gold standards” measurements
- Huge demand from EO and ML communities for better ground-truthing data
- Technology availability and adoption unevenly distributed across countries and constrained by analytical capacity

Against this backdrop …

ISWGHS

Positioning household surveys for the next decade

- ❑The document is being developed in response to the rapid changing landscape and the perceived need by member countries and DPs for a shared vision
- ❑Disclaimer: document is still being developed. The plan is to continue consultation to get inputs and share draft at next UNSC
- ❑9 priority areas, including enabling environment to accelerate realization of vision
- ❑Strike balance between foundational and frontier
  - x-country equity is strong consideration – with focus on LICs and LMICs where needs are greatest and both foundational and frontier features are weakest
- ❑Goal not to be comprehensive but ensure that key priorities are included, in light of recent developments and potential for medium-term impact
  - This webinar is part of our consultations and we hope to get your feedback

ISWGHS

9 priority areas

1. Enhance interoperability and integration of survey data
2. Improve sampling efficiency and coverage
3. Scale up adoption of improved methods and affordable technologies
4. Invest in capacity and research on CAPI, CATI, CAWI and mixed-mode
5. Understand and address non-response
6. Systematize collection, storage and use of meta/paradata
7. Expand capacity and use of machine learning and AI
8. Improve access, discoverability and dissemination of microdata
9. Foster a stronger operating environment

ISWGHS

Enhance interoperability/integration of survey data

- ❑Interoperability as one key attribute for (survey) data to generate value for development (Jolliffe et al, 2021)
  - Improve timeliness, cost-effectiveness, accuracy, and granularity of insights
  - Address issues of coverage and respondent burden
- ❑Interoperable surveys as a calibration/validation instrument for other data sources; examples in agriculture, poverty measurement and monitoring nutritional outcomes
  - Achieve economies of scale and scope through systematic collection of ground-truthing layers with multiple applications/uses
- ❑To ensure interoperability, must act at design stage (e.g., “poverty mapping”); it requires “interoperable institutions/actors”
- ❑To maximize benefits of interoperability, must promote data access while ensuring privacy
  - Need more research on spatial anonymization of public use datasets and comparative assessment of trade-offs between risk of disclosure and losses in empirical utility
  - Tailor to different users: spatially-anonymized public use datasets (w/ random offsets) versus confidential data accessed securely in data enclaves
- ❑Need to enhance statistical capacity in LICs on data integration and design, implementation and analysis of interoperable surveys

ISWGHS

Improve sampling efficiency and coverage

- ❑ Lacking/outdated censuses and under-coverage of important populations of interest major limitation of population-based survey frames
- ❑ Increase reliance on satellite imagery/geospatial data (e.g., HRSL) and multiple frames, particularly in conflict settings and to target hard-to-reach areas and population groups
- ❑ Increase piloting (and eventually scale up) of responsive/adaptive sampling techniques to meet specific needs
- ❑ COVID-19 phone surveys clearly show advantages of using recent nationally-representative F2F surveys as sampling frames
  - Requires a systematic approach to collecting contact info and tracing F2F survey respondents
  - Evidence on leveraging F2F survey data on bias adjustment at the household-level, but…
    - Limits to bias-adjustment at the individual-level (Brubaker et al., 2021)
    - Coverage concerns remain as a function of aging F2F survey samples used as frames
- ❑ Clear demand from countries to enhance statistical capacity and promote sustainable use of advanced sampling techniques

ISWGHS

Scale up adoption of improved methods and affordable technologies

- ❑ Non-classical measurement errors (NCME) in survey data have been shown to bias empirical analyses and policy recommendations
- ❑ Adopting improved survey methods, both "high-tech" sensors (e.g., handheld GPS for land area measurement, accelerometers for physical activity tracking) and "low-tech" objective measures (e.g., crop cutting for crop yield measurement), can help eliminate NCME
  - Cost may still be an issue. Multiple use/sharing? Imputation? Optimal size of sub-samples?
- ❑ High-frequency phone surveys in mixed-mode can reduce recall bias and improve timeliness/ frequency
  - There are, however, unanswered operational questions (e.g., assign cell phones? Set incentives and how much?)
- ❑ A business line on experimental statistics – *cum* capacity strengthening – can help streamline methodological research and development into NSO work program and help promote scaling up of validated methods

ISWGHS

Invest in capacity and research on CAPI, CATI, CAWI and mixed-mode surveys

- Rapid transition from PAPI to CAPI over the past decade; pandemic-induced move to CATI and CAWI in L/MICs; significant scope for building on this momentum to ensure more sustainable reliance on mixed-mode for recurrent events and crisis monitoring
- Capacity and infrastructure remain a constrain across NSOs in many countries (UNSD/WB, 2021)
- Need to invest in apps and tools for facilitating CATI/CAWI/mixed-mode surveys, including protocols for respondent selection, incentive provision and tools for data management
- Also, more randomized experiments and research needed on mode effect to understand data quality and comparability, likely to vary by variable

ISWGHS

Understand and address non-response

- With rapid urbanization and increasing income levels, non-response rates rising, even in LICs
  - Increasing complexity of instruments, waning trust in public institutions and privacy concerns making things worse
  - Shifts to new modes of data collection – e.g. phone, web – further accentuating problem

- More research and harmonization needed to guide ex-ante survey design choices to reduce response burden and increase trust/cooperation (e.g. contact protocols, questionnaire content and length, training protocols, …)

- Ability for ex-post corrections depends on availability of data, thus focus on design stage and metadata

ISWGHS

Systematize collection, storage and use of meta/ paradata

- CAPI/CATI/CAWI generating enormous amount of paradata which, if analyzed in real time, can be a game changer in fieldwork supervision and quality control.
- Possible uses include assignment of survey features based on previous responses and respondent's profile, gather info on contact burden, evaluate adoption of new data collection modes, etc.
- Also, systematic collection and use of metadata based on harmonized protocols extremely helpful for quality assurance as well as for treatment of non-response.
- Investments in protocols, apps and statistical tools for collection, storage and use/ sharing of meta/paradata should be prioritized.

ISWGHS

Expand capacity and use of ML and AI

- AI, ML and predictive analytics have great potential for virtually every step in the survey value chain, from questionnaire design to data collection and processing, to data use and dissemination
  - Predicting attrition in longitudinal surveys
  - Auto-coding of open-ended questions
  - Fast-tracking data editing and imputation
  - High-resolution mapping of development outcomes through data integration
  - Use of paradata for quality assurance

- Efforts still quite scattered and concentrated in HICs.

ISWGHS

Improve data access, discoverability and dissemination of microdata

- Address issue of underutilization of data – data use often constrained by data access and fitness-for-purpose of existing data
- Long-standing efforts – e.g. IHSN, Microdata Library, IPUMS … – but plagued by underinvestment and disincentives
- Data integration rest on improving access of both surveys (eg georeferencing) and other data sources (eg admin data)
- New types of data (e.g. georeferenced, rare populations, social media and other private data, …) create additional hurdles, raising additional privacy and ethical concerns; need new protocols and methods
- Capacity and IT infrastructure, particularly in LICs, remain constraint
- Suite of options by type of user/data? Institutional setting?

ISWGHS

Fostering a stronger operating environment at national, regional and global level

- ❑Strengthen engagement with data users and policy makers
- ❑Make research on survey methods easier to conduct and more "appealing" to researchers
- ❑Invest in ICT infrastructure for remote work, training, data collection …
- ❑Support (sub-regional?) hubs to strengthen capacity in use of new data sources and frontier applications/methods
- ❑Sustain financing and new financing models
- ❑Foster more coordinated and systemic approach to support NSOs
  - Sustain coordinated investments and research on global standards and data public goods
  - Support a stronger role of ISWGHS

ISWGHS

Some final thoughts …

- For household surveys to remain relevant and grow, we must strike the right balance between foundational and frontier work
- Alternative data sources provide an opportunity to add value to household surveys while also help increase spatial and temporal granularity
- The rapid diffusion of new, affordable technology a game changer to improve accuracy, coverage and use of household surveys
- The Covid-19 pandemic –potentially a major blow to the future of F2F surveys – has instead accelerated the process of modernization and innovation already in place
- More attention should go on quantifying benefits and communicating the value of survey data
- Rigorous methodological research addressing key measurement issues should be more systematic and systemic, as well as context-specific
- No technical solution will succeed and achieve scale without a proper enabling environment. Thus, the need for greater coordination and a more efficient use of available resources. The ISWGHS can help!

ISWGHS

THE INTER-SECRETARIAT
WORKING GROUP ON
HOUSEHOLD SURVEYS

Positioning Household Surveys for the Next Decade

Gero Carletto

gcarletto@worldbank.org

unstats.un.org/iswghs