

National Council on Family Relations Annual Conference, November 11, 2023 Orlando, FL

United Nations Department of Economic and Social Affairs (UNDESA)

Division for Inclusive Social Development

30th Anniversary of the International Year of the Family

North America United Nations Expert Group Meeting

**The Importance of Data in Understanding Demographic Changes and their Impact on
Families in the United States**

Mihaela Robila, PhD. CFLE
Professor, Human Development and Family Science
NCFR Representative at the United Nations (UN)
Interim Chair, Department of Family, Nutrition and Exercise Sciences
Queens College, City University of New York (CUNY)

The Importance of Data in Understanding Demographic Changes and their Impact on Families in the United States

Introduction

The rapid and significant growth of data availability and their uses place them at the forefront of the global agenda. The UN Secretary General's (2020a) *Data Strategy of the Secretary-General for Action by Everyone Everywhere with Insight, Impact and Integrity 2020-2022* proposed that data be considered and used as strategic assets to deliver sound policies and support services around the world. The outcomes pursued include "stronger decision making and thought-leadership, greater data access and sharing, improved governance and collaboration, robust data protection with respect for human rights, greater efficiency across our work, more transparency and accountability, and more relevant services for people and planet" (UN Secretary General, 2020a, p.2). The significant role of data is also recognized by *The Age of Digital Interdependence – Report of the UN Secretary-General's High-level Panel on Digital Cooperation* (UN Secretary General, 2019) and its recommendations resulted in the Secretary-General's *Roadmap for Digital Cooperation* (UN Secretary General, 2020b), in which the need to harness data for development is also emphasized. Increasing digitalization of the economy and society is changing the ways families interact.

The importance of harnessing the benefits of data has also been underlined at the global level as the 2030 Sustainable Development Goals (SDGs) were developed. The 2014 *A World That Counts: Mobilizing the Data Revolution for Sustainable Development* report established a new priority for the role of data in meeting the 2030 SDGs, recommending promoting innovation to fill data gaps, mobilizing resources to overcome inequalities between developed and developing countries, and coordination in using data to achieve sustainable development (IEAG, 2014).

Governments around the world recognize the increasingly importance of data for their societies' social and economic development and have developed national data governance strategies that ensure that institutions coordinate their efforts to foster gathering and using data according to each society's values, while protecting individuals' and families' rights over use of their information. Many countries have adopted initiatives to move towards integrated national systems (World Bank, 2021). For example, the *US Federal Data Strategy Framework* provides a 10-year vision on how the Government will foster the use of data to serve the public while preserving privacy and security and includes guidelines (e.g., exercise responsibility), goals (e.g., identify data needs) and actions (e.g., increase data skills) (US Government, 2020).

As data have value, there continues to be a tension between the expectation of open or free access and the private sector's drive to monetize them. Beyond the economic aspect, access to data needs to be considered in a broader context of human rights and security. Misuse of data leads to reduced public trust or could foster criminal activity. The lack of clear determination of borders across the digital world raises additional concerns related to ownership and access. Furthermore, access to data and data processing technologies is not uniform across the world, leading to deepening of economic inequalities, further compounded by heterogenous network connectivity (OECD, 2019; World Bank, 2021).

How data are generated also continues to evolve, in many cases requiring less and less human involvement or even awareness. Smartphones and related mobile devices and the ever-growing Internet of Things (IoT) technologies allow for large data collection on human activities (e.g., movements, purchases, online behaviors), but the reliance on such remote sensing can lead to a new type of inequality in access to devices or networks. The innovative potential of the new data streams calls for reflections on the new perspectives regarding data in the family and social science.

New Perspectives on Data

The new family data platforms and methodologies developments in the last decades provide opportunities to reflect on the different data infrastructures, including the traditional family national and international surveys, administrative data, as well as online and mobile data. When working in synergy and supported by novel processing approaches, the new and enhanced data sources provide opportunities to improve and diversify the richness of information (Callegaro, & Yang, 2018).

Surveys

International and national surveys have traditionally been a key component of the social data infrastructure, enabling research in family, social, and economic sciences and driving family and social policy making. Among them, household surveys have been a valuable tool for measurement of various socio-economic indicators (e.g., education, employment, poverty) (UN Statistics Division, 2020). However, continuous support for such surveys is increasingly challenging as the response rates continue to decrease despite diversification of interview modes. In their report on Positioning the Household Surveys for the Next Decade, the UN Inter-Secretariat Working Group on Household Surveys (UN ISWGHS) (2021) recommended “increasing the policy-relevance and use of household survey data to better inform policy and research; maximizing the efficiency and coverage of household surveys programs through coordination within the country and at the regional and international level” (p.3). Surveys remain key for measuring behaviors and attitudes at a depth that cannot be achieved through other means (Callegaro, & Yang, 2018).

The increase in globalization has increased the interest and opportunities to conduct multinational, and multiregional surveys comparing countries and regions on different issues including demographic changes, skills, or social opinions (Lyberg, Japac, & Tongur, 2019). Some of these surveys are funded by national budgets and some are supported by international organizations, such as the World Bank’s *Living Standards and Measurement Study*, USAID’s *Demographic and Health Surveys*, UNICEF’s *Multiple Indicator Cluster Survey*, and UN Women’s *Time Use Survey* (UN Statistics Divisions, 2020; UN Women, 2020). The *Living Standards and Measurement Study (LSMS)* has been supporting countries in designing data collecting mechanisms for key socioeconomic indicators (LSMS, 2020). While the surveys are country specific, the focus has been on measuring living conditions and the effectiveness of government policies and programs. LSMS makes the data openly accessible and attempts to aggregate them through common indicators. The LSMS’s work reflects the evolving state-of-

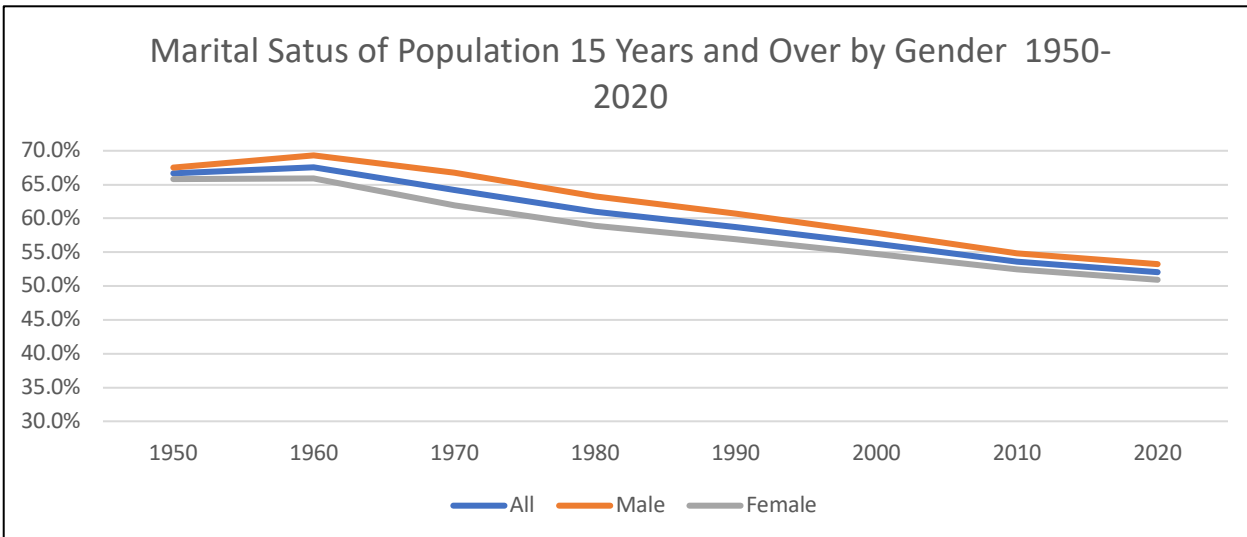
the-art in survey design and data gathering such as incorporation of online data collection and standardization of data analysis mechanisms.

The *Demographic and Health Surveys (DHS)* focus on health, family planning, gender, HIV, and nutrition indicators. Similar with LSMS, DHS supports survey design and collection, having enabled the development of hundreds of instruments in over 90 countries (DHS, 2019). To improve interoperability, DHS integrates multiple modules, and assists countries with data quality control and analysis, strengthening local methodological capacity. UNICEF's *Multiple Indicator Cluster Survey (MICS)* has been gathering information on children and women well-being. MICS was developed to support countries and respond to their data needs and has been adapted through the years to create a more useful data platform. More than 100 countries chose from the standard MICS questionnaire, modules that they adapt to fit their own data needs (UNICEF, 2021).

In the United States, the National Science Foundation has been funding large flagship surveys in social sciences such as the *General Social Survey*, and the *Panel Study of Income Dynamics*. The *General Social Science (GSS) Survey* has been gathering data on Americans' attitudes and behaviors related to inequalities, sociopolitical trends, religion, culture, health, social capital, and networks (Marsden, Smith, & Hout, 2020). The survey facilitates international comparison research through the *International Social Survey Program (ISSP)* that has now expanded to 60 countries. The *Panel Study of Income Dynamics (PSID)* is the world's longest running household longitudinal survey collecting data on economic and social wellbeing of the same families and their descendants (Johnson et.al., 2018). Topics covered include employment, income, wealth, poverty, consumer expenditures, health, marriage, child development, intergenerational relations, and neighborhood effects, among others. The *Current Population Survey (CPS)*, is sponsored by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), and is the primary source of labor force statistics for the US population.

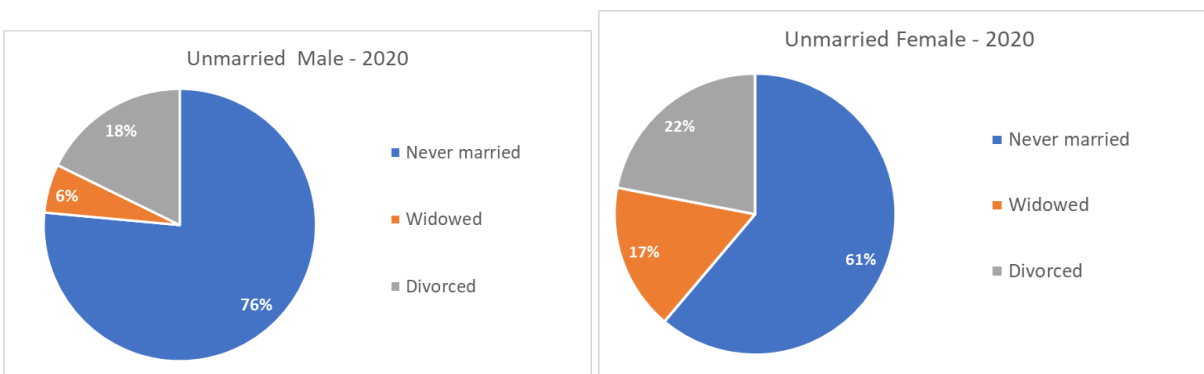
Data from the Current Population Survey (CPS) was used to examine the demographic changes in the United States, marriage and fertility rates, or aging trends. Marriage rates have decreased over the years from 66.6% in 1950 for to 52% in 2020, for both females (from 65.8% in 1950 to 50.9% in 2020) and males (from 67.5% to 53.2% respectively) (Graph 1). Figure 1 indicated that for the unmarried, 76% of men were never married, 18% were divorced and 6% widow, while 61% of women were never married, 22% divorced, and 17% widowed.

Graph 1. Marriage Rates



Data Source: U.S. Census Bureau, Decennial Censuses, 1950 to 1990, and Current Population Survey, Social and Economic Supplements

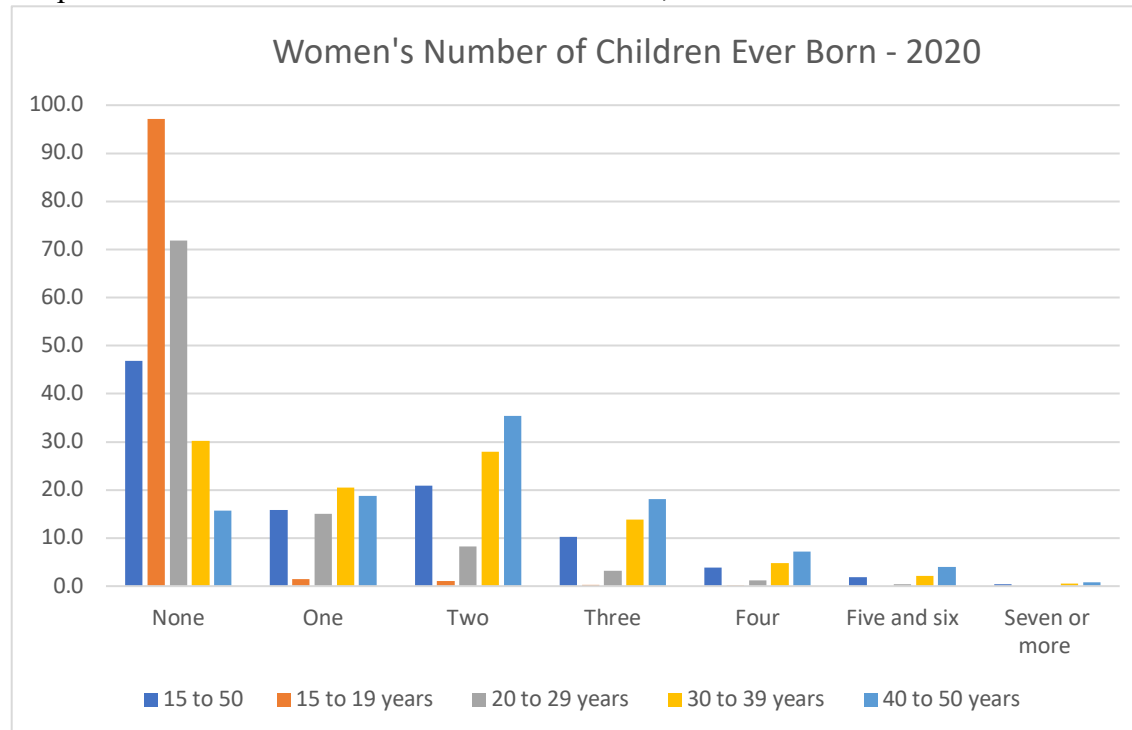
Figure 1. Unmarried Males and Females 2020



Data Source: U.S. Census Bureau, Decennial Censuses, 1950 to 1990, and Current Population Survey, Social and Economic Supplements

In terms of fertility rates, Graph 2 indicates women's number of children born and the age they give birth to their first child. For example, 72% of the 20–29-year-old women have no children, 15% have one, and 8% have 2 children, while 30% of the 30-39 year of women have no children, 21% have one and 18% have two children. In terms of fertility rates and marital status, 27% of 20–29-year-old women who are married have one child and 20% have two children, while 11% of the unmarried women have one children and 4.5% have two children (Table 1). The data also indicate that 21% of all married women (15 -50 years old) have one and 34% have two children, while 10% of unmarried women had 1 and 6% one child.

Graph 2. Women's Number of Children Ever Born, 2020



Data Source: U.S. Census Bureau, Current Population Survey, 2020

Table 1. Women's Number of Children Ever Born by Age and Marital Status: June 2020

All Women								
	All	None	One	Two	Three	Four	5 & 6	>=7
15 to 50	76,250	46.8	15.9	20.9	10.3	3.9	1.9	0.4
15 to 19 years	10,180	97.2	1.5	1.0	0.3	0.1	0.0	0.0
20 to 29 years	21,760	71.9	15.0	8.3	3.2	1.2	0.4	0.0
30 to 39 years	22,010	30.2	20.6	28.0	13.9	4.8	2.1	0.5
40 to 50 years	22,310	15.7	18.7	35.4	18.1	7.2	4.0	0.8
Married		None	One	Two	Three	Four	5 & 6	>=7
15 to 50	40,340	18.2	21.2	34.2	16.6	6.3	2.9	0.6
15 to 19 years	142	55.5	27.8	16.8	0.0	0.0	0.0	0.0
20 to 29 years	5,638	43.4	26.5	19.3	7.9	2.4	0.6	0.1
30 to 39 years	15,547	18.4	22.0	34.3	16.8	5.6	2.4	0.6
40 to 50 years	19,015	10.4	18.8	38.7	19.2	8.0	4.1	0.8
Never Married		None	One	Two	Three	Four	5 & 6	>=7
15 to 50	35,920	78.9	10.0	6.1	3.1	1.1	0.7	0.1
15 to 19 years	10,040	97.8	1.1	0.7	0.3	0.1	0.0	0.0
20 to 29 years	16,118	81.9	11.0	4.5	1.6	0.8	0.3	0.0
30 to 39 years	6,457	58.7	16.9	13.1	6.9	2.6	1.5	0.3
40 to 50 years	3,302	46.7	18.5	16.4	11.6	2.7	3.2	0.9

Source: U.S. Census Bureau, Current Population Survey, 2020

The data from the National Household Education Surveys Program (Table 2) indicated that more children under 1 years old are in relative care than those in center-based care, while for children who are 1-2 and 3-5 the numbers of center-based childcare is higher than those with relative care, and the numbers are similar for boys and girls. More single parents use relative care than two parent families. Interestingly, families with the highest income are the least to use relative care, while more families with lower incomes use relative childcare arrangements.

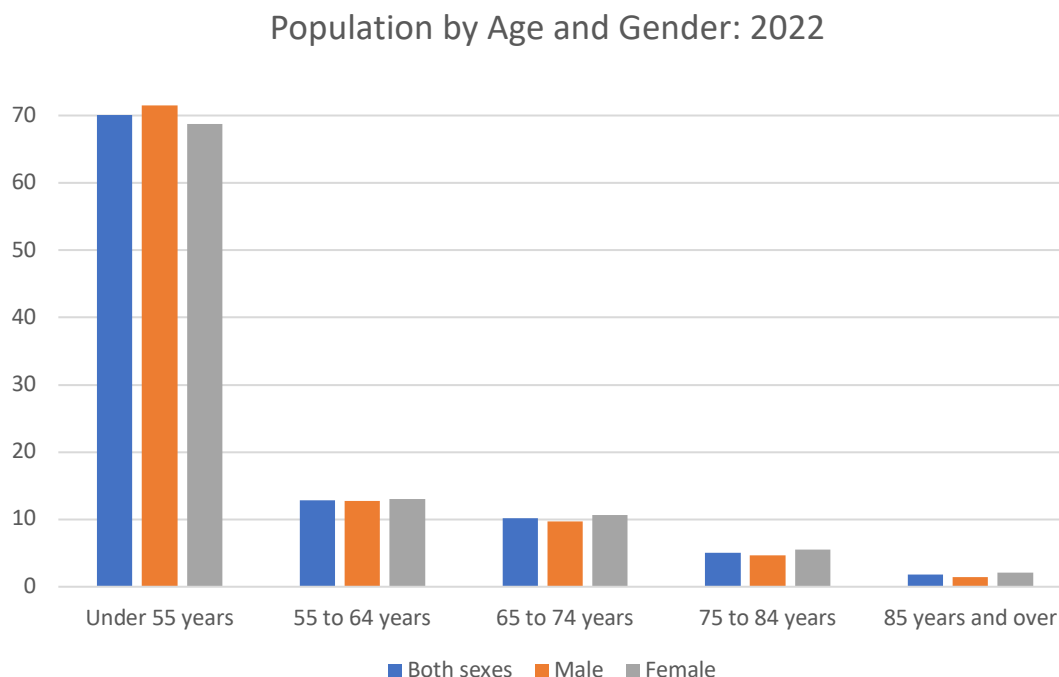
Table 2. Child Care Arrangements

Characteristic	Number of children (thousands)	At least one weekly nonparental care arrangement	Type			nonparental care arrangement
			Relative care	Nonrelative care	Center-based care	
Total	21,195	59	38	20	62	41
Child's age						
< 1 year	4,621	42	58	26	32	58
1–2 years	8,425	55	44	25	47	45
3–5 years	8,149	74	26	14	83	26
Child's gender						
Male	10,992	59	38	19	64	41
Female	10,203	60	38	20	61	40
Race/ethnicity						
White	10,420	61	34	23	65	39
Black	2,706	63	45	13	59	37
Hispanic	5,424	56	44	16	57	44
Asian/ Pacific Is.	1,181	55	35	12	67	45
Other race	1,463	59	32	23	62	41
Family type						
Two parents	17,105	58	35	20	63	42
One parent	4,089	65	50	17	60	35
Labor force						
Two-parent						
Both full time	6,401	86	37	25	59	14
One full one p.	2,860	64	41	19	56	36
One full one n.	6,315	31	19	10	83	69
Single-parent						
Full time	2,136	75	48	20	63	25
Part time	716	78	55	12	51	22
Not working	936	36	55	71	60	64
Household income						
\$20,000 or less	2,401	51	43	17	64	49
\$20,001–\$50,000	5,063	46	45	14	59	54
\$50,001–\$75,000	3,659	55	46	22	50	45
\$75,001–\$100k	2,849	58	42	18	59	42
> \$100,001	7,223	74	29	22	69	26

Note. Data source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Program Participation Survey of the 2019 National Household Education Surveys Program (ECPN-NHES:2019)

In United States, Census and CPS data indicated that in 2022, 70% of people were 55 or younger, 13% were 55 to 64 years old, 5% were 75 to 85 years old. For the 85 and older, 2.1% were females and 1.4% were males (Graph 3).

Graph 3. Older Population by Age and Gender



Data Source: U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, 2022

Data from the CDC National Center for Health Statistics National Post-Acute and Long-Term Care Study, 2020-2021 (Table 3) indicated that 88% of the users of Residential Care Community are White, 5% are Black and 3% Hispanic, while higher percentages of minority groups are using the Adult Day Services Centers (22% Hispanics, 18% Asian and 15% Black, 40% White). More than twice as many women (69%) than men (31%) are using the Residential Care Community. Half of the users of the Residential Care Community are 85 years old or older, and most of these facilities (82%) are for profit and only 1 % are owned by the government.

Table. 3. Elder Care Services

	Residential Care Community	Adult Day Services Centers
Capacity		
Average licensed maximum capacity or number of beds	39	71
Total licensed maximum capacity or total number of beds	1,197,600	294,000
Number of users	818,800	237,400
Ethnicity/race		
Hispanic	2.93%	21.83%
Non-Hispanic American Indian or Alaska Native	0.53%	0.65%
Non-Hispanic Asian	1.91%	18.47%
Non-Hispanic Black	4.94%	15.01%
Non-Hispanic Native Hawaiian or other Pacific Islander	0.46%	0.72%
Non-Hispanic Other race category or two or more races	1.25%	3.23%
Non-Hispanic White	87.98%	40.10%
Gender		
Female	69.46%	56.68%
Male	30.54%	43.32%
Age		
Under 65	5.83%	36.71%
65 and over	94.17%	63.29%
65-74	13.40%	23.71%
75-84	30.89%	24.94%
85 and over	49.88%	14.64%
Medicaid for payment	17.51%	61.82%
Ownership		
For-profit	81.88%	45.53%
Government and other	1.06%	3.68%
Nonprofit	17.06%	50.80%

Note. Data Source: CDC National Center for Health Statistics National Post-Acute and Long Term Care Study, 2020-2021.

The increase in the elderly population determined an increase in the need for health care and social assistance services. The data from the Service Annual Survey (SAS) on nationwide estimates of revenue and expenses of different service industries that serve the nation's elderly indicate an increase in revenues from 2013 to 2021, for Home Health Care Services by 56%, for Continuing Care Retirement Communities and Assisted Living by 37% and for Services for the Elderly and Persons with Disabilities by 71%. (Table 4).

Table 4. Estimated Revenue by Tax Status for Long-term Care Providers

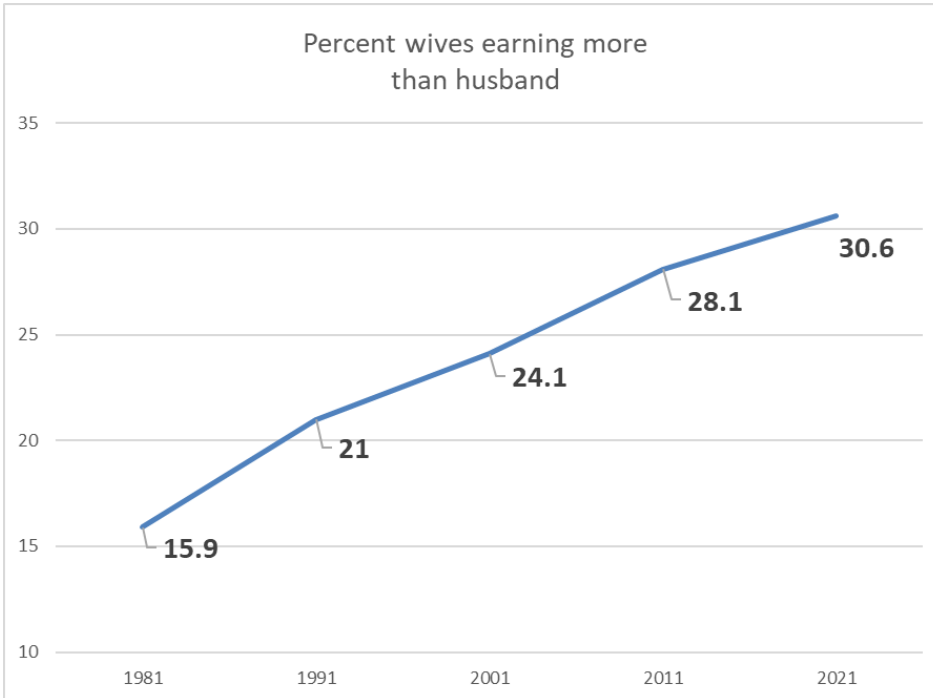
Facility Type	2013	2014	2018	2019	2020	2021	Overall change (2013-2021)
Home Health Care Services	68,286	71,648	89,453	97,263	100,551	106,696	56.2%
Nursing Care Facilities (Skilled Nursing Facilities)	110,833	112,989	120,191	123,830	128,617	128,329	15.8%
Continuing Care Retirement Communities and Assisted Living Facilities for the Elderly	53,634	56,389	68,082	72,610	73,306	73,403	36.9%
Assisted Living Facilities for the Elderly	25,062	26,256	31,305	33,253	33,430	34,019	35.7%
Services for the Elderly and Persons with Disabilities	35,441	38,740	52,057	55,028	57,446	60,618	71.0%

Note. Data source: U.S. Census Bureau. 2021 Service Annual Survey and administrative data. Estimates are in millions of dollars

Earnings and employment have also varied over the years. Graph 4 indicated an increase in the percentages of wives whose earnings are greater than husbands' earnings from 16% in 1981 to 31% in 2021. There has been a steeper increase between 1981 (16%) and 1991 (21%), and a much smaller one between 2011(28.1%) and 2021 (30.6%).

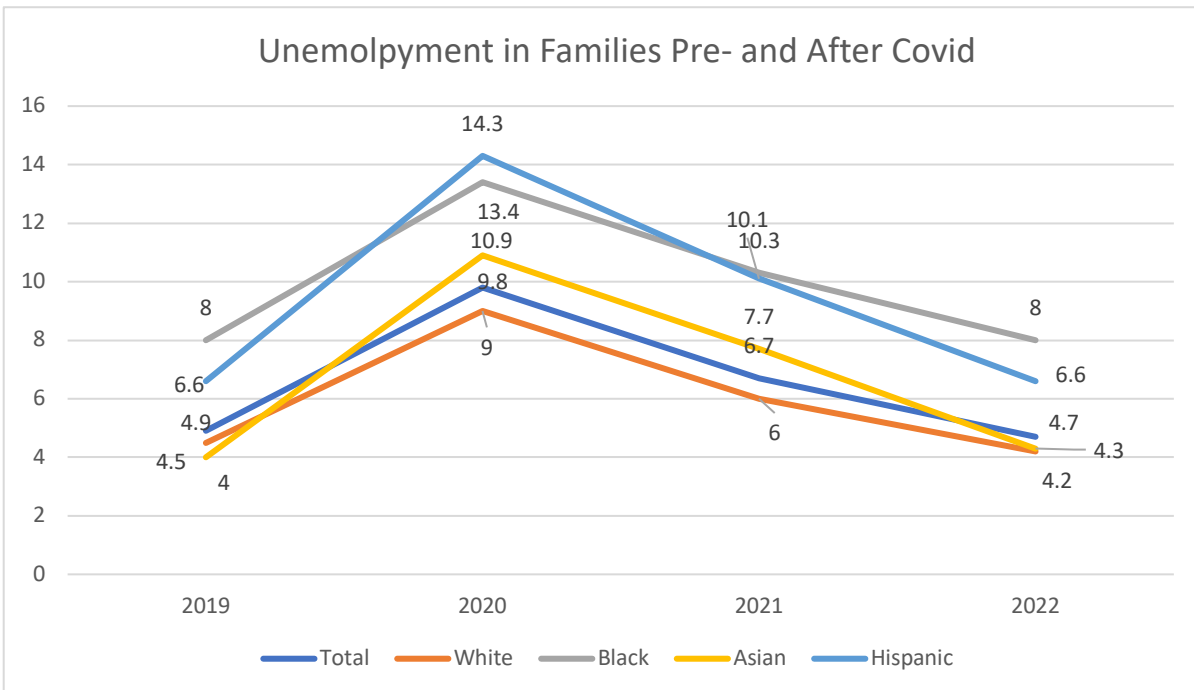
COVID pandemic had a profound impact on employment and earnings. Graph 5 illustrated the rates in unemployment for families pre-and post-pandemic. For all families it varied from 4.9% in 2019, to 9.8% in 2020, to 4.7% in 2022, while for black families it varied from 8% in 2019, to 13.4% in 2020 to 8% in 2022, and for Hispanic families, from 6.6% in 2019, 14.3% in 2020, and 6% in 2022.

Graph 4. Married-Couple Families with Wives' Earnings Greater than Husbands' Earnings



Data Source: Source: U.S. Census Bureau, Current Population Survey, 1982 to 2022 Annual Social and Economic Supplements (CPS ASEC)

Graph 5. Unemployment Pre-and Post Covid



Data Source: US Bureau of Labor Statistics

While each of these surveys continues to evolve in response to changes in communication modalities, they also illustrate the need for innovation to maintain their leadership in the field. Besides updates introduced to surveys themselves, innovation could come from increased interoperability across the surveys, and their integration with other data sources such as administrative data and online or mobile big data.

Administrative Data

Information on social and behavioral human activity has been gathered as part of administrative data from governmental and non-governmental records. In the United States, administrative data include governmental data at the federal (e.g., income tax records, Social Security) and state level (e.g., unemployment, housing, educations records), and non-government data (e.g., credit card transactions) (Groves, & Schoeller, 2018). The increased digitalization of the society has resulted in both an increase in the amount of data collected and in a lowering of the barriers of access to data.

By being continuously collected and refreshed, administrative data are particularly valuable in reflecting up-to-date societal trends. Because of this, administrative data play an important role in policy making at local and national levels. Such data can also be used in research and practice. As administrative data collection is already part of organizational data infrastructure, reusing it is cost efficient. However, as such data are generated by various entities, significant challenges remain in integration, anonymization, availability and coverage. Administrative data are limited to people using the different social programs and services that generate them, are constrained to the time period that the person participates in the program and are less structured than the systematically collected data (Groves, & Schoeller, 2018).

Research on family self-sufficiency and wellbeing, needs and the effectiveness of interventions has used and linked administrative data from public benefits programs such as TANF, SNAP, Medicaid, child care subsidies, disability, housing assistance, unemployment insurance, and public education and from public programs systems such as child welfare, homeless shelters, criminal justice (Cuccaro-Alamina et.al., 2021; Wiegand, & Goerge 2019). For example, Goerge, & Wiegand (2019) examined Illinois families' use of multiple public-sector services from five state agencies, to understand the receipt of child welfare, mental health, and substance abuse services. Thus, administrative data on Supplemental Nutrition Assistance Program (SNAP) participation (from the Illinois Department of Human Services), on child welfare service records (from the Illinois Department on Child and Family Services), on incarceration records (from Illinois Department of Corrections and the Illinois Department of Juvenile Justice), and Medicaid claims for treatment for mental health and substance abuse, and the results indicated that 23% of the 502,165 families included in the study were multisystem families, and accounted for 86% of the funding for health, mental health, criminal and juvenile, and child welfare needs and another 34% of families received services in one of the five areas and accounted for the remaining funds (Goerge, & Wiegand 2019).

Enhancements in both administrative and survey data collection pipelines, and development of new protection strategies in response to increased privacy threats resulting from aggregation of multiple sources are needed (NASEM, 2017). Along with the various administrative data, the

large amounts of online and mobile data could significantly contribute to the social sciences. These potentials need to be explored to capitalize on the advantages and limit the possible risks involved.

Online and Mobile Big Data

Several converging trends have led to the generation and use of data resulting from online activity and / or use of mobile devices. First, changes in human activities, ushered by the Digital Age have led to an extraordinary amount of information being generated by individuals' and families' online activities as well as through the use of various devices (smartphones, GPS, smartcards, etc.). Second, computing technology advances have made it possible to collect and store such data at ever finer granularity, and new algorithmic solutions have been developed to process them. Finally, significant economic value has been attached to such data through monetization by digital information companies.

The large amount of data generated by online activities, internet searches, social media, sensors, mobile phones, has fostered opportunities to study individuals' and families' behavior and social interactions. In particular, social media platforms such as Twitter and Facebook have already been used extensively in family and social science research (e.g., McCormick, et. al., 2017). Many of these platforms are regularly used by individuals or families to find information and support with different questions. For example, Teague and Shatte, (2018) used social media to explore the transition to fatherhood on web-based communities, by mapping the discussion topics on two forums for expectant and current fathers from the social media site Reddit and identified clusters of topics such as pregnancy milestones, first-time fathers, practices, and challenges. Online social support systems can thus be designed for parents, to reach and engage fathers in childrearing.

As Big Data are characterized by missing or inconsistent values and convenience samples, applying traditional statistics approaches is often not feasible. Moreover, from the point of view of social and behavioral science research, not all Big Data are useful. There are also ethical, access and representativeness problems. Much of these data are unstructured, meaning they are not organized compared to the systematically collected and assembled data which are easy to use. In order to be analyzed, Big Data require complex methodologies of mathematical models and artificial intelligence techniques (e.g., machine learning, natural language processing). Thus, these new methodologies necessitate proper data analytics and computational education and skills and provide opportunities for interdisciplinary collaborations between social scientists and computer scientists (Robila, & Robila, 2020).

Surveys and online and mobile data can serve complementary functions, with Big Data providing opportunities for improving surveys, and some survey frameworks and methods being applicable to big data (Hill, et.al., 2021). The potential of the available new data sources is higher when they are linked to traditional survey data and thus data from various sources (e.g., private sector data, internet data/social media) can be processed and integrated to support effective policymaking decisions (NASEM, 2017).

Opportunities and Challenges

While surveys have been providing important information to various stakeholders, including the government, economy, and health and education systems, their role has been challenged by declining respondents' participation, increasing costs, and the competition from other data sources (Miller, 2017). Surveys have been important components of national statistical systems in many countries but are now being questioned in terms of their usefulness and adaptability (UN Statistics Division, 2020). UN Women underlines the importance of data disaggregation by sex, disability status, location and other statuses to understand the intersecting inequalities, gender equality and women empowerment, in order to ensure leaving-no-one-behind (UN Women, 2020). Integration must also be actively pursued as part of a coordinated strategy to prevent data overlap and gaps, and leading to a more cost effective and sustainable data infrastructure (Sabiti, Anderson, & Wozniak, 2021; UNICEF 2019).

Ongoing developments to address the challenges encountered in survey research include new data collection strategies and innovative approaches to integrating survey and non-survey data. Recent decades have seen a continuous growth in online survey data collection across the world, as well as a diversification of the non-survey digital data collection. In the United States there are several mixed methods national surveys, including the *Panel Study of Income Dynamics*, and the *Current Population Survey* which is administered by the Census Bureau. Integrating surveys with other data sources can bring benefits such as producing more granular and timely data and serving as a source of validation for other data (UN Statistics Division, 2020).

Surveys have developed rigorous mechanisms to ensure respondents' anonymity and confidentiality during data collection, processing and results dissemination, and similar privacy, ethical, and legal protocols need to be used for administrative and online and mobile Big Data collection and management (Callegaro, & Yang, 2018). Using social media platforms for data analysis presents important opportunities as well as new risks. Given the open nature of such platforms, ensuring data privacy and confidentiality may require new ethics guidelines (Berman, Powell, & Garcia Harranz, 2018). One common problem with social media platforms is network privacy: people's networks may be involved and included without their consent and knowledge even if the primary participant consents. When social media is used to engage more vulnerable populations, such as children, even more attention needs to be provided to ensure children's well-being and to assess the potential benefits and risks.

The *2021 World Development Report* "calls for a new social contract for data that enables the use and reuse of data to create economic and social *value*, promotes *equitable* opportunities to benefit from data, and fosters citizens' *trust* that they will not be harmed by misuse of the data they provide" (World Bank, 2021, p. xi). The report proposes a conceptual framework that *engages governmental and non-governmental organizations, the private sector, academia, and the civil society in building data pathways that enable each partner in achieving their specific mandates*.

In the United States there have been efforts to enhance data protection, such as the *Federal Data Strategy Framework* which includes principles explicitly referring to ethical governance and

responsibility of federal data access and use (US Government, 2020). Moreover, policies have also been enacted by Congress and States, such as California, Florida, Illinois, among others, to also provide protection to citizens' privacy (NCLS, 2022).

Four themes are predominant in terms of the future of data (see Figure 2). First, the next generation infrastructure will be one that integrates different types of data from surveys, administrative records, online and mobile data to obtain more comprehensive information that could be effectively used in policymaking and program and service development and implementation. While surveys will continue to play an important role in quality assurance and filling data gaps, administrative and online data can provide important and timely contributions. Second, there are multiple entities that will serve as data sources and compete for data provision and use. Novel collaboration mechanisms will also emerge, supporting both interdisciplinary research efforts and data sharing. Third, with further technological advances, it is expected that other sources of data will appear and there will be a more dynamic processes in choosing different types of data over others. Fourth, and equally important, is the need for new skills development for the advanced analytical and methodological knowledge necessary to manage large and diverse data streams, thus requiring support of learning and workforce development programs. The integration of new data streams has the potential to enable innovation in social science research by opening new directions of inquiry.

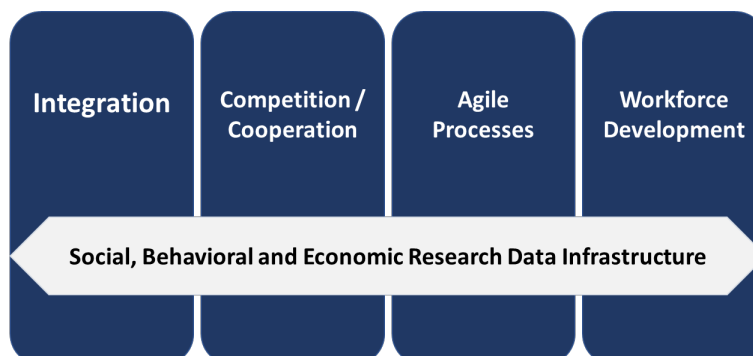


Figure 2. Data Infrastructure

Conclusions

A future family, social and behavioral science data infrastructure will rely on the integration of a variety of open data from different sources (e.g., survey, administrative, online and mobile data) generated by public and private organizations. Public data, such as the ones generated as part of the design, execution, and evaluation of programs and services, together with commercially generated ones, collected and curated by the private sector, when integrated with traditional survey data can open new transformative understandings of the human, social and economic development (World Bank, 2021). Data need to be collected and managed so that they are compatible and integrated with other data sets, are available for secondary data analysis, and for replications of studies are open to the research community. Large integrated data systems present also challenges such as confidentiality and security breaches (Lane, 2018).

High quality data are needed for evidence-based effective policymaking (Reamer, & Lane, 2018). Data in all sectors of the economic and social development are necessary to monitor and ensure achieving the 2030 Sustainable Development Goals, to examine the progress as well as the gaps that remain and still need to be addressed.

The COVID-19 pandemic has accelerated the process of digital transformation as individuals, governments, and society at large had to find ways to adapt and continue their activities. During this time, new methods of data collection have been adopted in many countries (e.g., web surveys) and alternative data sources have been used more, and they are likely to continue in the future (UN ISWGHS, 2021). Interagency and international coordination for data is also necessary, as fostering data sharing and analysis is an important matter, high on the national and global agenda. High quality data has the potential to reduce costs and enabling global scientific discoveries to improve people's lives around the globe.

Recommendations

- Designing national and international data governance frameworks to promote the development of effective data coordination mechanisms to ensure societal data needs assessments, data gathering and use while respecting data privacy and ethics.
- Enhancement in both administrative and survey data collection pipelines, development of new protection strategies in response to increased privacy threats resulting from aggregation of multiple sources
- Ensure data availability and integration key in making informed decision at the local level; national statistics should also include mechanisms for generation of disaggregated regional data
- Ensure data integration, as part of a coordinated strategy, to prevent data overlap and gaps, and promote cost effective and sustainable data
- Engages governmental and non-governmental organizations, the private sector, academia, and the civil society in building data pathways that enable each partner in achieving their specific mandates

References

- Berman, G., Powell, J., & Garcia Harranz, M. (2018). *Ethical Considerations When Using Social Media for Evidence Generation*. Innocenti Discussion Paper 2018-01, UNICEF Office of Research - Innocenti, Florence.
- Burke-Garcia, A., Edwards, B., & Yan, T. (2021). The Future Is Now: How Surveys Can Harness Social Media to Address Twenty-first Century Challenges. In Hill et al. (Eds.), *Big Data Meets Survey Science: A Collection of Innovative Methods* (p. 63-99). Hoboken, NJ: John Wiley & Sons.
- Callegaro, M., & Yang, Y. (2018). The Role of Surveys in the Era of “Big Data”. In D.L. Vannette, J.A. Krosnick (eds.), *The Palgrave Handbook of Survey Research* (p.175-192) Center for Disease Control (CDC) National Center for Health Statistics (NCHS). (2021) National Post-acute and Long-term Care Study. <https://www.cdc.gov/nchs/npals/webtables/overview.htm>
- Cuccaro-Alamina, S., Eastmana, A.L., Fousta, R., McCroskeya, J., Nghiema, H.T., Putnam Hornstein, E. (2021). Strategies for constructing household and family units with linked administrative records. *Children and Youth Services Review*, 120. <https://doi.org/10.1016/j.chilyouth.2020.105706>
- Demographic and Health Surveys (DHS). (2019). *Demographic and Health Surveys Brochure*. USAID.
- Goerge, R. M., & Wiegand, E. R. (2019). Understanding vulnerable families in multiple service systems. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(2), 86–104.
- Groves, R.M., & Schoeller, G.J. (2018). Use of Administrative Records in Evidence Based Policymaking. *ANNALS of AAPSS*, 678, 71-80.
- Hill, C. A., Biemer, P. P., Buskirk, T.D., Japac, L., Kirchner, A., Kolenikov, S., & Lyberg, L.E. (2021). Introduction. In Hill et al. (EDS.), *Big Data Meets Survey Science: A Collection of Innovative Methods* (pp.1-9). Hoboken, NJ: John Wiley & Sons.
- Independent Expert Advisory Group (IEAG) (2014). *A World That Counts: Mobilizing the Data Revolution for Sustainable Development*. Indep. Expert Advisory Group Secretariat.
- Johnson, D.S., McGonagle, K.A., Freedman, V.A., & Sastry, N. (2018). Fifty Years of the Panel Study of Income Dynamics: Past, Present and Future. *ANNALS AAPSS*, 680, 9-28.
- Lane, J. (2018). Building an Infrastructure to Support the Use of Government Admin. Data for program performance and Social Science Research. *ANNALS, AAPSS*, 675,240-52.
- Lyberg, L., Japac, L., & Tongur, C. (2019). Prevailing issues and the future of comparative surveys. In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts* (3MC) (pp. 1055–1082). John Wiley & Sons, Inc.
- Living Standards Measurement Study (LSMS). (2020). *Living Standards Measurement Study Brochure*. World Bank Developing Data Group.
- Marsden, P.V., Smith, T.W., & Hout, M. (2020). Tracking US Social Change over a Half-Century: The General Social Survey at Fifty. *Annual Review of Sociology*, 46, 109-134.
- McCormick, T. H., Lee, H., Caesare, N., Shojale, A., & Spiro, E.S. (2017). Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. *Sociological Methods & Research*, 46(3), 90-421.
- Miller, P. (2017). Is There a Future for Surveys? *Public Opinion Quarterly*, 81, 205-212.

- National Academies of Sciences, Engineering, and Medicine (NASEM). (2017). *The Value of Social, Behavioral, and Economic Sciences to National Priorities: A Report for the National Science Foundation*. Washington, DC: The National Academies Press.
- National Conference on State Legislatures (NCSL) (2022). *Privacy Protections on State Constitutions*.
- OECD. (2019). *Data in the Digital Age*. OECD Going Digital Policy Note, OECD, Paris,
- Reamer, A., & Lane, J. (2018). A Roadmap to a National Data Infrastructure for Evidenced-Based policymaking. *ANNALS, AAPSS*, 675, 28-35.
- Robila, M. & Robila, S. A. (2020). Applications of Artificial Intelligence Methodologies to Behavioral and Social Sciences. *J. of Child and Family Studies*, 29(10), 2954-2966.
- Sabiti, B., Anderson, B. & Wozniak, S. (2021). *The Data Side of Leaving No One Behind: Lessons from Landscaping*. Development Initiatives.
- Teague, S. J., & Shatte, A. B. (2018). Exploring the Transition to Fatherhood: Feasibility Study Using Social Media and Machine Learning. *JMIR Pediatrics and Parenting*, 1(2), e12371.
- UNICEF. (2021). *Multiple Indicator Cluster Survey*.
- UNICEF. (2019). *Key asks for 2030 SDG Voluntary National Reviews: SDG 9*.
- UNICEF. (2017). *Data For Children Strategic Framework*.
- UN Inter-Secretariat Working Group on Household Surveys (UN ISWGHS). (2021). *Positioning the Household Surveys for the Next Decade*. 52nd Session of the UN Statistical Commission.
- UN Secretary-General. (2020a). *Data Strategy of the Secretary General for Action by Everyone Everywhere with Insight, Impact and Integrity 2020-2022*.
- UN Secretary-General. (2020b). Road Map for Digital Cooperation: Implementation of the Rec. of the High-level Panel on Digital Cooperation Report of the Secretary-General.
- UN Secretary General (2019). *The Age of Digital Interdependence – Report of the UN Secretary-General’s High-level Panel on Digital Cooperation*.
- UN Statistics Division. (2020). *Household Surveys in a Changing Data Landscape; Challenges, Opportunities and an Agenda for the Future*. Side ev. 51st Session UN Stat Commission.
- UN Women. (2020). *UN Women Statement at the 51st Session of the UN Statistical Commission: Better Data Better Lives*. New York, NY: United Nations.
- U.S. Census Bureau (2021). Annual Service Survey.
<https://www.census.gov/data/tables/2021/econ/services/sas-naics.html>
- U.S. Department of Education, National Center for Education Statistics, Early Childhood Program Participation Survey of the 2019 National Household Education Surveys Program (ECP-P-NHES:2019).
- US Government. (2020). *Federal Data Strategy Framework: Mission, Principles, Practices and actions*.
- Wiegand, E. R. & Goerge R. M. (2019). *Using and linking administrative datasets for family self-sufficiency research*. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.
- World Bank (2021). *A World Bank Flagship Report: Data for Better Lives*. Washington, DC.