

DATA PAPER

The Adolescent Data Hub: The Largest Catalog of Open Access Data on Adolescents Living in Low- and Middle-Income Countries

Corinne White, Christina Misunas, Erica Chuang, Midanna de Almada, Stephanie Psaki, and Thoai D. Ngo

The Adolescent Data Hub (ADH) is the first and largest data catalog specifically developed to focus on open access data on adolescents in low- and middle-income countries (LMICs). Developed by the Population Council's Girl Innovation, Research, and Learning (GIRL) Center, and launched in August 2018, the ADH has grown to include more than 750 data sources that fit the inclusion criteria of (1) self-reported data from females and/or males between ages 10 and 24 years; (2) one or more rounds of data collected in year 2000 or later; (3) data collected in one or more LMICs; (4) data are publicly available for download and use. A dynamic resource, the ADH is regularly updated to include new datasets that meet these criteria. The ADH facilitates access to available data on adolescents for researchers attempting to answer important questions related to the lives of adolescents and for donors and policymakers eager to identify gaps in existing data to inform their future investments.

PURPOSE AND VALUE OF THE CATALOG

Adolescent Data Needs

The world's more than 1.8 billion adolescents (aged 10–24) face a complex set of challenges that are being met with increased funding for interventions, research, and advocacy across health and development sectors (United Nations 2019; UNICEF 2016; Li et al. 2018). Members of the international community have designed and implemented global initiatives such as the Global Strategy for Women's, Children's and Adolescents' Health (2016–2030) as well as country-level policies to improve conditions for adolescents and help countries meet the Sustainable Development Goals (WHO 2015). However, a changing landscape of demographic and epidemiological factors has impeded

Corinne White, Erica Chuang, Stephanie Psaki, and Thoai D. Ngo, Population Council, Poverty, Gender, and Youth Program, New York, NY 10017. Christina Misunas, Erica Chuang, Stephanie Psaki, and Thoai D. Ngo, Population Council, Girl Innovation, Research, and Learning (GIRL) Center, New York, NY 10017. Email: cmisunas@popcouncil.org. Midanna de Almada, Cooper/Smith, Washington, DC 20006.

progress on many fronts. Since 1990, population growth has been the highest in low-resource countries with the worst adolescent health profiles based on communicable disease, maternal health, and nutritional indicators; these countries also face lower rates of secondary school completion and higher rates of adolescent fertility, child marriage, and unemployment (Azzopardi et al. 2019). Researchers and policymakers have also pointed to gaps in available data as a constraint to informed decision-making, underscoring the necessity of both expanding and leveraging the evidence base of research on adolescents (Azzopardi et al. 2019).

In recent years, there has been a growing recognition of the need to coordinate both ongoing and future research, with the goal of building intersectoral, multicomponent interventions that promote positive adolescent health and well-being (Patton et al. 2016). The Data2X initiative, one example of such coordination, serves as a technical and advocacy platform by fostering and leveraging partnerships to improve the availability and accuracy of gender disaggregated data. To date, the initiative has mapped broad gaps in data across five domains of women's empowerment and offered recommendations for future data collection efforts, specifically on social determinants of adolescent girls' health, socially excluded girls, transitions from education to the workforce, and international comparability of education data (Buvinic, Furst-Nichols, and Koolwal 2014). However, without a current and comprehensive assessment of available data, initiatives like Data2X might struggle to accurately gauge the most pressing gaps and to efficiently drive resources toward new data collection.

To address this challenge and help inform future solutions, the Population Council's Girl Innovation, Research, and Learning (GIRL) Center created the Adolescent Data Hub (ADH). To our knowledge, the ADH is the largest online catalog of open-access survey data on adolescents and young people living in low- and middle-income countries (LMICs). The GIRL Center designed the ADH to facilitate secondary analyses by connecting researchers to existing data and to provide insights into geographic and topic area data availability. The ADH provides a platform for data sharing at scale for the community of researchers, donors, and implementers working on adolescent issues.

Open Data

Research transparency efforts in the social sciences are motivated by scholarship that has highlighted a lack of replicability of psychology and economics studies; shaped by ethical research principles and collaborative scientific norms; and influenced by trends in transparency in other fields, including the biomedical sciences (Open Science Collaboration 2015; Christensen and Miguel 2018; Wallach, Boyack, and Ioannidis 2018). Across disciplines, researchers have developed strategies to alleviate the problems that opaque processes can exacerbate (Nosek et al. 2015; Miguel et al. 2014). Organizations, social networks, and study protocol registries have facilitated collaboration between researchers and have shared information about field experiments prior to implementation.

Additionally, research funders, including the Bill & Melinda Gates Foundation (BMGF), the United Kingdom Department for International Development (DFID), the U.S. National Institutes of Health (NIH), the United States Agency for International Development (USAID), and Wellcome Trust, as well as various government entities and intergovernmental

agencies, have established data-sharing policies and developed open data portals (BMGF, n.d.; DFID 2013; NIH 2017; USAID, n.d.; Wellcome Trust, n.d.). Several academic publishers have introduced tiered data-sharing policies that either encourage or mandate the practice for their authors (Elsevier n.d.; Springer Nature n.d.; Wiley n.d.). Likewise, some multilateral and research organizations and government agencies have created open data portals that range from narrow to general in scope, such as the Harvard Dataverse, Institute for Health Metrics and Evaluation's (IHME) Global Health Data Exchange, and the World Bank Open Data portal (The President & Fellows of Harvard College 2019; IHME, n.d.; World Bank, n.d.).

Data sharing operationalizes aspects of research transparency and helps avoid waste in the research pipeline (Chalmers and Glasziou 2009). Data-sharing behaviors appear to have increased during the past decade, but many researchers still consider their own lack of expertise on data management and curation a roadblock to consistent implementation (Tenopir et al. 2015). Some researchers might also have concerns about their ability to publish using secondary data; however, an analysis of bibliometric data on a neuroscience open-data initiative found that secondary analyses of shared data were published in moderate- to high-impact journals at frequencies comparable to primary data analyses (Milham et al. 2018). Trends toward increased data sharing, especially when supported by institutional structure, provide an opportunity to accelerate collaboration across disciplines, drive more efficient use of resources, and ultimately facilitate the generation of more evidence for decision-makers.

The Adolescent Data Hub: Goals and benefits

The design of the ADH enables researchers to assess the availability of existing data and therefore increases opportunities for comparative and cross-sectoral research. The ADH facilitates secondary data analysis by reducing the time and effort needed to identify open access data on a topic of interest. It can be a particularly beneficial resource for early-career researchers and those in resource-constrained settings, who might lack a network of international researchers from whom to learn about available data, and who might not have the donor or institutional support needed for primary data collection.

Some of the benefits of building and maintaining a comprehensive catalog of open-access data sources include (1) enabling further analysis of existing data to build on the evidence base; (2) encouraging more collaboration between researchers; (3) promoting research transparency, integrity, and reproducibility; (4) ensuring resources are not wasted on duplicative work; and (5) expanding researchers' access to data sources. For implementers and decision-makers, the ADH can help to identify geographic and thematic gaps in data availability on adolescents and consequently to inform programming, policy, and investment in new data collection or analysis.

METHODS

Inclusion Criteria

The following criteria were used for including data sources in the catalog:

1. Self-reported quantitative data from females and/or males between ages 10 and 24 years¹;
2. One or more rounds of data collected in year 2000 or later²;
3. Data collected in one or more LMICs³;
4. Data are publicly available for download and use.⁴

We have currently restricted the scope of the catalog to LMICs in order to focus on regions and countries where the bulk of this population lives and where needs of adolescents are potentially most acute (Azzopardi et al. 2019). We also used the year 2000 as a threshold to coincide with the start of the Millennium Development Goals. These criteria were established at the onset of compilation and might be subject to revision based on user feedback and emerging research priorities.

Data Source Identification and Categorization

To identify available open-access data sources, we initially undertook the following processes and have repeated them quarterly:

1. A scan of existing databases (of open-access data sources) using keywords related to inclusion criteria;
2. A broader Internet search using keywords related to inclusion criteria;
3. Direct and indirect outreach to researchers studying adolescents in LMICs.

Within the catalog, we further categorized data sources by geographic region and country; study design (observational or experimental); data structure (cross-sectional or longitudinal); background of study respondents (female and/or male; age category 10–14, 15–19, and/or 20–24); and thematic topics covered. Nationally representative data sources were also noted as such.

To identify a broad list of thematic topics, we assessed (1) a subset of open-access data sources and their respective questionnaires and (2) priorities emerging from adolescent-focused research. We later identified sub-topics based on more specific themes commonly emerging from the questionnaires. We anticipate expanding the lists of topics and subtopics based on feedback from users and evolving research priorities (see Table 1).

DESCRIPTION OF CONTENT

The GIRL Center launched the ADH in August 2018 with 128 data sources. As of November 2019, the catalog includes 755 publicly available data sources. The ADH includes a variety of study types, including major study series (e.g., Demographic and Health Surveys), single

1 Sources with data on adolescents reported by an intermediary source, such as a parent or teacher, were excluded. The WHO's definition of adolescents (aged 10–19) and young people (aged 20–24) were used. <https://apps.who.int/adolescent/second-decade/section2/page1/recognizing-adolescence.html>

2 As we build the catalog, we made a strategic decision to focus on the most recent data but might later expand to include earlier data sources. The year 2000 cutoff may be subject to change in the future.

3 The World Bank's categorization of low- and middle-income countries was used (World Bank 2019).

4 Datasets are housed on external websites. Levels of access are predetermined by their respective principal investigators.

TABLE 1 Among the sources included in Adolescent Data Hub, the percent with data on each topic and subtopic, by region

Topic	Total	Region					
		South Asia	Europe and Central Asia	Middle East and North Africa	Sub-Saharan Africa	Latin America and Caribbean	East Asia and Pacific
<i>Community engagement</i>	28	31	21	40	27	33	24
<i>Crime</i>	28	29	29	19	25	41	25
<i>Economics</i>	72	71	74	72	71	73	70
Employment	52	53	47	53	56	52	46
Savings	13	16	17	21	11	9	8
<i>Education</i>	99	98	100	100	99	98	99
Grade attainment/achievement	98	98	100	100	99	95	98
Learning	25	33	22	18	31	17	25
School quality	2	7	0	3	3	2	2
<i>Family/household structure</i>	98	93	100	96	97	100	98
<i>Gender attitudes</i>	49	53	45	53	53	47	42
Endorsement of violence against women	37	31	36	31	44	41	24
Gender role beliefs	36	35	29	44	38	35	33
<i>Health care access and utilization</i>	51	55	37	38	63	50	41
<i>HIV/AIDS and STIs</i>	49	53	32	41	64	36	45
Knowledge	43	50	31	41	50	34	44
Testing	35	35	30	18	48	26	30
Treatment	8	12	6	3	11	3	7
<i>Marriage</i>	62	71	64	62	64	56	59
<i>Mental health</i>	43	43	54	38	30	61	47
Life satisfaction	21	12	45	18	11	33	15
<i>Migration and mobility</i>	33	28	42	37	31	35	26
<i>Physical health</i>	47	55	50	50	43	45	55
Anthropometry	20	29	20	24	14	19	31
Nutrition	31	38	25	29	30	30	37
Self-reported health	30	31	37	27	29	29	26
<i>Reproductive health</i>	64	64	66	44	67	69	61
Family planning/contraceptives	39	47	40	25	48	26	3
Female genital mutilation/cutting	6	0	0	7	15	0	0
Fertility	56	60	56	43	60	55	48
Maternal health	8	16	6	6	8	5	9
Menstrual hygiene	2	5	0	3	4	0	2
Sexual behavior	40	33	36	9	52	37	42
<i>Social networks</i>	23	38	16	24	17	23	38
<i>Subjective expectations</i>	24	22	22	22	25	29	22
<i>Time use</i>	33	36	28	31	33	31	36
<i>Tobacco/alcohol/drug use</i>	58	60	56	49	56	60	67
<i>Violence</i>	41	40	32	25	42	54	41
Emotional/psychological violence	27	28	31	24	23	32	29
Physical violence	39	40	32	24	39	54	40
Sexual violence	30	26	21	24	28	46	30
Total number of data sources (n)	755	58	118	68	282	129	100

country cross-sectional studies, single or multicountry longitudinal studies, and subnational or population-specific cross-sectional and longitudinal studies. The ADH also includes several Population Council-led studies that were publicly unavailable until 2018, such as the Malawi Schooling and Adolescent Study, the Adolescent Girls Empowerment Program (Zambia), and the Tathmini GBV Study (Tanzania) (Hewett and Mensch 2015; Austrian 2018a; Settergren et al. 2018).

TABLE 2 Number of data sources that are part of survey series

Series Title	n
AIDS Indicator Survey	7
Balkan Epidemiological Study on Child Abuse and Neglect	8
Barometer Surveys	81
Demographic and Health Surveys	68
Demographic and Health Special Surveys	5
European Social Survey	6
Global Early Adolescent Survey	3
Global School-Based Student Health Survey	72
Global Youth Tobacco Survey	132
Health Behaviour in School-Aged Children Survey	10
International Crime Victimization Survey	24
Multiple Indicator Cluster Survey	112
National Survey of Adolescents	4
PACARDO: Data on Drug Use and Behavior in School-Aged Children and Teenagers in Panama, Central America, and the Dominican Republic, 1999–2000	7
Performance Monitoring and Accountability 2020 (PMA2020)	11
Safe and Smart Savings	2
Skills Toward Employment and Productivity Measurement Household Survey	13
Violence Against Children and Youth Survey	8
WHO Study on Global AGEing and Adult Health (SAGE)	6
World Health Survey	44
World Values Survey	57
Young Lives	4
Total number of data sources (n)	684

Series

Data from survey series, such as Demographic and Health Surveys (DHS) or the Violence Against Children Surveys (VACS), represent approximately 92 percent ($n = 684$) of all sources featured in the ADH⁵ (see Table 2 for full list) (ICF, n.d.; CDC, n.d.). Survey series allow for international and sometimes subnational comparisons. The use of standardized questionnaires, survey instruments, and implementation guidelines in many survey series often enables researchers to compare findings within and/or across countries and subpopulations. In some instances, the frequency of data collection (e.g., every 2 to 5 years) enables analyses of trends across and/or within cohorts.

Longitudinal Studies

Less than 10 percent of the data sources included in the ADH are longitudinal.⁶ Longitudinal data are valuable to researchers interested in understanding causative factors and the timing of major life transitions during adolescence. However, despite its unique value, longitudinal data collection requires more time and often longer-term funding from donors. Examples of longitudinal studies in the ADH include the Indonesia Family Life Survey, Young Lives (Ethiopia, India, Peru, Vietnam) and Safe and Smart Savings (Kenya, Uganda) (RAND Corporation, n.d.; Young Lives 2019; Austrian 2018b).

5 All descriptions of ADH composition are accurate as of November 2019 and are subject to change as ADH evolves.

6 All sources are categorized based on whether they have longitudinal or cross-sectional data. There are some studies flagged as cross-sectional that have panel data (repeated cross sections); for all cross-sectional sources, we note that number of survey rounds for each source.

Experimental Studies

Approximately 3 percent of data sources included in the ADH are from experimental studies. While the main findings from experiments have likely been published, secondary analysis of these data can allow researchers to further understand how and why an intervention worked, and/or investigate other questions that were beyond the scope of the original study. These experiments span multiple dimensions of adolescent health and well-being, including a cash transfer intervention to encourage school attendance in Malawi (Schooling, Income, and Health Risk Impact Evaluation Survey) and the relationship between HIV prevention efforts and exposure to sexual risk in Cameroon (Risk Information and Adolescent Sexual Behavior) (Baird, McIntosh, and Özler 2013; Dupas, Huillery, and Seban 2017).

Other Studies

While 80 percent of the data sources included in the ADH are considered nationally representative, subnational data can provide more nuanced insight into a topic or subpopulation that might not be captured in larger, nationally representative studies. For example, the Urban Life among Youth in Kisumu Project in Kenya employed a novel survey instrument called a Relationship History Calendar that collected detailed retrospective information on young people's sexual relationship histories (Luke, Clark, and Zulu n.d.). Similarly, the Rural Youth Survey in Tunisia collected data on the unique economic and livelihood opportunities for young people in rural areas (Institut National de la Statistique 2015).

Major Topics and Gaps

Data sources are categorized by region, country, age, and sex of respondents and key thematic areas pertinent to adolescent health and development issues. As shown in Table 1, among all data sources in the ADH, regional representation is as follows: East Asia and Pacific (13 percent), Europe and Central Asia (16 percent), Latin America and the Caribbean (17 percent), Middle East and North Africa (9 percent), Sub-Saharan Africa (37 percent), and South Asia (8 percent).

Within regions, country-level representation varies. For example, Kenya, ($n = 21$), Ghana ($n = 17$), and South Africa ($n = 16$) are strongly represented among data sources in Sub-Saharan Africa ($n = 282$), whereas Angola, Central African Republic, and Eritrea each have only two data sources that meet our inclusion criteria. Notably, countries currently undergoing conflict, such as South Sudan ($n = 2$) and Syrian Arab Republic ($n = 3$), have disproportionately fewer studies. In addition, the distribution of both experimental and longitudinal evidence varies by region. We did not identify any experimental studies in Latin America and the Caribbean or Europe and Central Asia; similarly, only three longitudinal surveys in Latin America and the Caribbean and just one in Europe and Central Asia met our inclusion criteria.

Data availability also varies based on age of the respondents covered: 39 percent of data sources contain information on very young adolescents (aged 10–14), 99 percent on older adolescents (aged 15–19), and 62 percent on young adults (aged 20–24) (Table 3). Gaps in coverage emerge for data on very young adolescents; topics less commonly covered among

TABLE 3 Among the sources included in Adolescent Data Hub, the percent with data on each topic and sub-topic, by age group

Topic	Adolescent age group		
	10–14	15–19	20–24
<i>Community engagement</i>	15	28	44
<i>Crime</i>	7	29	36
<i>Economics</i>	66	72	74
Employment	18	52	71
Savings	6	13	19
<i>Education</i>	99	99	99
Grade attainment/achievement	98	99	98
Learning	7	25	40
School quality	5	2	3
<i>Family/household structure</i>	24	68	92
<i>Gender attitudes</i>	16	49	76
Endorsement of violence against women	8	38	60
Gender role beliefs	14	35	54
<i>Health care access and utilization</i>	16	51	70
<i>HIV/AIDS and STIs</i>	32	49	55
Knowledge	31	43	45
Testing	12	35	44
Treatment	3	8	10
<i>Marriage</i>	18	62	88
<i>Mental health</i>	38	43	42
Life satisfaction	4	21	32
<i>Migration and mobility</i>	14	33	50
<i>Physical health</i>	44	48	48
Anthropometry	44	20	5
Nutrition	36	31	22
Self-reported health	13	30	38
<i>Reproductive health</i>	38	64	78
Family planning/contraceptives	18	39	49
Female genital mutilation/cutting	1	6	10
Fertility	17	56	77
Maternal health	1	8	3
Menstrual hygiene	1	2	3
Sexual behavior	33	41	40
<i>Social networks</i>	41	23	18
<i>Subjective expectations</i>	9	24	38
<i>Time use</i>	35	33	24
<i>Tobacco/alcohol/drug use</i>	86	58	36
<i>Violence</i>	44	41	44
Emotional/psychological violence	40	27	24
Physical violence	42	39	42
Sexual violence	37	30	29
Total number of data sources (n)	295	747	465

very young adolescents include crime, gender attitudes, life satisfaction and subjective expectations, literacy and numeracy, migration and mobility, and certain aspects of reproductive health, particularly female genital mutilation/cutting (FGM/C).

Education

Nearly all data sources in the ADH have data on school enrollment and grade attainment. Only a quarter of sources have data on learning outcomes, which are often collected through survey instruments that assess literacy and/or numeracy directly. Of the sources with data on learning outcomes, almost 80 percent are from series such as DHS or UNICEF's Multiple Indicator Cluster Survey (MICS) (ICF, n.d.; UNICEF 2019). Less than 3 percent of all sources have data on school quality; however, this topic is more commonly covered among

surveys of younger adolescents (5 percent) and among surveys conducted in South Asia (7 percent).

HIV/AIDS and STIs

About half of all sources in the ADH have data on HIV/AIDS and STIs. Forty percent include data on respondents' knowledge of transmission, and approximately 35 percent have data on testing, either through self-reported information or biomarker test results.⁷ However, less than 10 percent of data sources have information on whether respondents are or have formerly received treatment, which limits our understanding of respondents' health-seeking behavior and access to services.

Reproductive Health

Approximately two-thirds of the sources in the ADH have data on adolescent reproductive health. Compared to other world regions, Middle East and North Africa appears to have fewer data sources on reproductive health, specifically regarding family planning/contraceptive use and sexual behavior. Overall, commonly covered subtopics include fertility (55 percent), sexual behavior (40 percent), and family planning/contraceptive use (39 percent). Many of these data sources originate from the DHS and MICS series, which include women's questionnaires, specifically documentation of birth histories (ICF, n.d.; UNICEF 2019). Less than 3 percent of all data sources contain information on menstrual hygiene and about 8 percent cover maternal health. No open-access data sources on menstrual hygiene were identified in Latin America and the Caribbean or Europe and Central Asia. Approximately 6 percent of all data sources include data on FGM/C, although it is a practice that has been documented in 30 countries, mainly in Africa, the Middle East, and Asia, (UNICEF 2013). Sexual violence is more commonly covered among data sources in Latin America and the Caribbean (46 percent), compared to the other regions.

USAGE

The ADH's interactive design enables users to browse the full inventory of sources included in the catalog as well as search for relevant sources using specific terms. Through various viewing formats, the ADH also allows users to visualize data availability within and across regions based on select criteria (Figure 1). Users can choose to view data sources in alphabetical or chronological order and can further organize them based on relevant region, study design, data structure, and backgrounds of respondents (age category or sex).

As shown in Figure 2, a profile of each data source is provided with information on the study's purpose, design, participants, and topics covered as well as the investigators, organizations, and funders involved. Links to the respective study websites are provided for users to download data and access other relevant materials. Additional country-level statistics on key adolescent-focused indicators are presented alongside each data source to provide country context.

⁷ We do not tag whether HIV/STI testing information was collected by self-report or biomarkers.

FIGURE 1 Example of grid view functionality—grouping sources by region (Sub-Saharan Africa), with search narrowed by topic (social networks)

The screenshot shows the Adolescent Data Hub interface. At the top, it displays '177 Datasets', '6 Regions', and '89 Countries'. Below this, there are filters for 'View: Map | Grid | List', 'Color: Region | Sex | Study Design | Type of Data', 'Group: No Group | Region | Age Category | Sex', and 'Order: Most Recent | Alphabetical'. The main content is a grid of dataset cards for the 'Sub-Saharan Africa' region, filtered by the topic 'Social Networks'. The grid contains 14 dataset cards, each with a title, description, and a color-coded background.

FIGURE 2 Example of how a data source is profiled in the ADH

Adolescent Girls Empowerment Programme (AGEP)			Request Data
<p>The Adolescent Girls Empowerment Program is a randomized evaluation testing the effectiveness of a set of interventions aimed at improving vulnerable girls' social, health, and economic resources so that they can stay in school longer; avoid early marriage; delay sexual activity; and prevent unintended pregnancy, HIV, and other STIs.</p> <p>Investigators: Karen Austrian (Population Council), Paul C. Hewett (Population Council)</p> <p>Funders: UK Department for International Development (DFID)</p> <p>Series: Not applicable</p> <p>Notes: Age range during first survey round.</p>	<p>Adolescents: 2,500-4,999</p> <p>Sex: Females</p> <p>Age Range: 10-19</p> <p>Study Design: Experimental Longitudinal</p> <p>Type of Data:</p> <p>Time Period: 2017; 2016; 2015; 2014; 2013</p> <p>Areas covered: Provinces: Lusaka; Central; Copperbelt; North-Western</p>	<p>Zambia; Sub-Saharan Africa</p> <p>Population:^a 18.1 million</p> <p>% Adolescents:^a 3400%</p> <p>GNI per capita:^b \$3,582</p> <p>MEAN YEARS OF SCHOOLING^b</p> <p>Female: 6.7</p> <p>Male: 7.5</p> <p>LIFE EXPECTANCY AT BIRTH^b </p> <p>Female: 66.4</p> <p>Male: 60.5</p> <p>Adolescent birth rate:^b </p> <p>120.1</p>	<p>Request Data</p>
<p>Topics: Education - Grade attainment/attendance, Learning, School quality; Physical Health - Self-reported health, Nutrition; Mental Health; Reproductive Health - Sexual behavior, Fertility, Family planning/contraception; HIV/AIDS and STIs - Testing, Knowledge; Marriage; Gender Attitudes/Beliefs - Gender role beliefs, Endorsement of violence against women, Marriage attitudes; Violence - Sexual violence, Physical violence, Emotional/psychological violence; Migration and Mobility; Social Networks; Economics - Employment, Savings; Demographic Characteristics - Urban and Rural; Family/Household Structure; Young Adolescents; Very Young Adolescents; Older Adolescents</p>			<p>Study Website</p>

VALUE OF THE CATALOG

The ADH advances research transparency by providing a platform to share and access publicly available data on adolescents living in LMICs in service of a more collaborative and efficient research environment. This catalog is unique in its focus on adolescents, a demo-

graphic group that has received growing attention and investment from the global health and development community (Sawyer 2018; WHO 2015; Li et al. 2018).

The ADH can serve as a valuable resource to the field for the following reasons:

- Data sources span a range of disciplines, including economics, education, health, and other development sectors. In compiling seemingly disparate sources into a comprehensive catalog, the ADH facilitates secondary analysis of existing data and potentially sparks unique, cross-sectoral research on adolescents.
- It assists in identifying experimental or intervention-based studies with robust data that researchers can ideally use to demonstrate what works, what does not work, and what has not yet been tested.
- Longitudinal evidence available in the ADH allows researchers to investigate causality and understand the sequence of events and changes experienced by adolescents over time. By providing a catalog of existing longitudinal data sources, the ADH affords researchers with opportunities to assess whether changes occurring over time are consistent across populations.
- Finally, the wealth of survey series that can be accessed through the ADH enables researchers to compare patterns, trends, and prevalence across diverse contexts, often using standardized instruments and processes.

LIMITATIONS

Although the GIRL Center works to regularly identify new open-access data sources to ensure the catalog is comprehensive and up to date, we are ultimately unable to determine the catalog's level of completeness. Additional desk research and outreach to researchers will presumably result in the identification and inclusion of other open-access data sources on adolescents. More broadly, many existing datasets on adolescents are not yet publicly available, particularly those collected most recently. Consequently, our assessment of data availability and gaps may be biased due to missing data. As donors and journals increasingly require data sharing, we expect the ADH to even more closely reflect the full universe of existing data on adolescents. The list of thematic topics and subtopics used to categorize data sources is limited and potentially excludes other important, context-dependent topics. In addition to categorizing sources based on the ages and sexes of respondents covered, other important demographic information (i.e., religion, ethnicity, and socioeconomic status) have likely been collected and might be useful for some researchers. The ADH does not include qualitative data sources nor does it distinguish sources that collected qualitative data in addition to quantitative. Thus far, we have excluded programmatic quantitative data due to a general lack of institutional review board approval for many programs that are not associated with research studies. We compiled information on each data source from relevant websites and other publications; while we attempted to resolve any reporting discrepancies, in most cases we did not contact study investigators to confirm the accuracy of the information in the available materials. Finally, due to the number of sources included in the catalog, we do not attempt to assess the data quality of each source.

REFERENCES

- Austrian, Karen. 2018a. "Adolescent Girls Empowerment Program (AGEP)." <https://doi.org/10.7910/DVN/CFIUC6>, Harvard Dataverse, V1.
- . 2018b. "Safe and Smart Savings." <https://doi.org/10.7910/DVN/SOW4HP>, Harvard Dataverse, V2.
- Azzopardi, Peter S., Stephen J.C. Hearps, Kate L. Francis, Elissa C. Kennedy, Ali H. Mokdad, Nicholas J. Kassebaum, Stephen Lim, et al. 2019. "Progress in Adolescent Health and Wellbeing: Tracking 12 Headline Indicators for 195 Countries and Territories, 1990–2016." *The Lancet* 393(10176): 1101–1118. [https://doi.org/10.1016/2FS0140-6736\(18\)32427-9](https://doi.org/10.1016/2FS0140-6736(18)32427-9)
- Baird, Sarah, Craig McIntosh and Berk Özler. 2013. "Malawi - Schooling, Income, and Health Risk Impact Evaluation Household Survey 2007–2008, Round I (Baseline). Ref. MWI_2007_SIHR_v01_M_v01_A_PUF." Dataset downloaded from <https://microdata.worldbank.org/index.php/catalog/1005>
- Bill & Melinda Gates Foundation (BMGF). n.d. "Open Access Policy." Accessed May 2, 2019. <https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>
- Buvinic, Mayra, Rebecca Furst-Nichols, and Gayatri Koolwal. 2014. "Mapping Gender Data Gaps." Data2x. Accessed November 25, 2019. https://data2x.org/wp-content/uploads/2019/05/Data2X_MappingGenderDataGaps_FullReport.pdf
- CDC. n.d. "Violence against Children Surveys: Towards a Violence-Free Generation." <https://www.cdc.gov/violenceprevention/childabuseandneglect/vacs/index.html>
- Chalmers, Iain, and Paul Glasziou. 2009. "Avoidable Waste in the Production and Reporting of Research Evidence." *Obstetrics & Gynecology* 114(6): 1341–1345. <https://doi.org/10.1097/aog.0b013e3181c3020d>.
- Christensen, Garret, and Edward Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature* 56(3): 920–980.
- DFID. 2013. "DFID Research Open and Enhanced Access Policy." Accessed November 25, 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/181176/DFIDResearch-Open-and-Enhanced-Access-Policy.pdf
- Dupas, Pascaline, Elise Huillery and Juliette Seban. 2017. "Risk information, risk salience, and adolescent sexual behavior: Experimental evidence from Cameroon." <https://doi.org/10.7910/DVN/MLVGY9>, Harvard Dataverse, V1.
- Elsevier. n.d. "Research Data Guidelines." Accessed May 2, 2019. <https://www.elsevier.com/authors/author-resources/research-data/data-guidelines>
- HBSC. n.d. "Health Behaviour in School-Aged Children: World Health Organization Collaborative Cross-National Survey." <https://www.hbsc.org/>
- Hewett, Paul C. and Barbara Mensch. 2015. "Malawi Schooling and Adolescent Study Data Sets." <https://www.popcouncil.org/research/malawi-schooling-and-adolescent-study-data-sets>
- ICF. n.d. "The DHS Program: Demographic and Health Surveys." <https://dhsprogram.com/data>
- IHME. n.d. "Global Health Data Exchange." Accessed May 2, 2019. <http://ghdx.healthdata.org/>
- Institut National de la Statistique. 2015. "Tunisia – Rural Youth Survey 2012." http://microdata.worldbank.org/index.php/catalog/2333/related_materials
- Li, Zhihui, Mingqiang Li, George C. Patton, and Chunling Lu. 2018. "Global Development Assistance for Adolescent Health From 2003 to 2015." *JAMA Network Open* 1(4): e181072. <https://doi.org/10.1001/jamanetworkopen.2018.1072>.
- Luke, Nancy, Shelley Clark and Eliya Zulu. n.d. "Life History Data: Urban Life among Youth in Kisumu Project (ULYKP)." McGill University. <https://www.mcgill.ca/lifehistoriesandhealth/data/lifehistorydata>
- Miguel, Edward, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M. Esterling, Alan Gerber, Rachel Glennerster, et al. 2014. "Promoting Transparency in Social Science Research." *Science* 343(6166): 30–31.
- Milham, Michael P., R. Cameron Craddock, Jake J. Son, Michael Fleischmann, Jon Clucas, Helen Xu, Bonhwang Koo et al. 2018. "Assessment of the Impact of Shared Brain Imaging Data on the Scientific Literature." *Nature Communications* 9, 2818. <https://doi.org/10.1038/s41467-018-04976-1>.
- NIH. 2017. "NIH Data Sharing Policy." Last updated April 17, 2017. https://grants.nih.gov/grants/policy/data_sharing/
- Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck et al. 2015. "Promoting an Open Research Culture." *Science* 348(6242): 1422–1425.

- Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science*, 349(6251): p.aac4716. <https://doi.org/10.1126/science.aac4716>.
- Patton, George C., Susan M. Sawyer, John S. Santelli, David A. Ross, Rima Afifi, Nicholas B. Allen, Monika Arora et al. 2016. "Our Future: A *Lancet* Commission on Adolescent Health and Wellbeing." *The Lancet* 387(10036): 2423–2478. [https://doi.org/10.1016/S0140-6736\(16\)00579-1](https://doi.org/10.1016/S0140-6736(16)00579-1).
- RAND Corporation. n.d. "The Indonesia Family Life Survey (IFLS)." <https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html>
- Sawyer, Susan M. 2018. "Towards a Global Alliance for Adolescent Health." *Journal of Adolescent Health* 63(1): 7–8. <https://doi.org/10.1016/j.jadohealth.2018.05.006>
- Settegren, Susan K., Stella Mujaya, Wasima Rida, et al. 2018. "Tathmini GBV study: Evaluation of Comprehensive Gender-Based Violence Programming Delivered through the HIV Program Platform in Tanzania." <https://doi.org/10.7910/DVN/05DPTL>, *Harvard Dataverse*, V1.
- Springer Nature. n.d. "Research Data Policy Types." Accessed May 2, 2019. <https://www.springernature.com/gp/authors/research-data-policy/data-policy-types/12327096>
- Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLoS One* 10(8): e0134826. <https://doi.org/10.1371/journal.pone.0134826>
- The Presidents and Fellows of Harvard College. 2019. *Harvard Dataverse*. <https://dataverse.harvard.edu/>
- UNICEF. 2013. "Female Genital Mutilation/Cutting: A Statistical Overview and Exploration of the Dynamics of Change." New York: UNICEF.
- . 2016. "Adolescent Demographics." <https://data.unicef.org/topic/adolescents/demographics/>
- . 2019. "Multiple Indicator Cluster Surveys (MICS)." <http://mics.unicef.org/>
- USAID. n.d. "Development Data Library (DDL)." Accessed May 2, 2019. <https://data.usaid.gov/>
- United Nations, Department of Economic and Social Affairs, Population Division. 2019. *World Population Prospects 2019*, Online Edition.
- Wallach, Joshua D., Kevin W. Boyack, and John P.A. Ioannidis. 2018. "Reproducible Research Practices, Transparency, and Open Access Data in the Biomedical Literature, 2015–2017." *PLoS Biology* 16(11): e2006930. <https://doi.org/10.1371/journal.pbio.2006930>
- Wellcome Trust. n.d. "Open Research." Accessed May 2, 2019. <https://wellcome.ac.uk/what-we-do/our-work/open-research>.
- WHO. 2015. "The Global Strategy for Women's, Children's and Adolescents' Health (2016–2030)." Accessed November 14, 2019. <https://www.who.int/life-course/partners/global-strategy/ewec-globalstrategyreport-200915.pdf?ua=1>
- Wiley. n.d. "Wiley's Data Sharing Policies." Accessed May 2, 2019. <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>
- World Bank. 2019. "World Bank Country and Lending Groups." *World Bank Country and Lending Groups – World Bank Data Help Desk*. <datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>.
- World Bank. n.d. "World Bank Open Data." Accessed May 2, 2019. <https://data.worldbank.org/>
- Young Lives. 2019. "University of Oxford Department of International Development." <https://www.younglives.org.uk/>

ACKNOWLEDGMENTS

The authors thank all of the research teams who have provided open access to data included in the ADH. The Ford Foundation under Grant 131039, the William and Flora Hewlett Foundation under Grant 2017–5126, the John D. and Catherine T. MacArthur Foundation under Grant 17-109327-000-POP, and the David and Lucile Packard Foundation under Grant 2017-66689 provided financial support for the creation and maintenance of the online catalog. Sophia Chae, Christina Tse, and Meredith Kozak were instrumental in the creation and launch of the ADH. Two-N developed and designed the website.