EXECUTIVE SUMMARY

Recommendations on the
Ethical Use of Novel HIV Data & Analytics

Working Group on the Ethical Use of Novel HIV Data and Analytics

May 2024

Funded by the Bill and Melinda Gates Foundation

ethics lab
Georgetown University
Center for Innovation in Global Health
COOPER/SMITH
Bill & Melinda Gates foundation

WORKING GROUP MEMBERS

Janet Tatenda Bhila
Zimbabwe Young Positives (ZY+)

Aleny Couto, MD
Ministry of Health, STI and HIV/AIDS, Mozambique

Shona Dalal, PhD
WHO, Dept. of Global HIV, Hepatitis, and STIs

Will Fleisher, PhD
Georgetown University, Center for Digital Ethics

Jen Gennai
Google, Responsible Innovation

Rayid Ghani
Carnegie Mellon University, Machine Learning Dept. and College of Information Systems and Public Policy

Peter Godfrey-Faussett
London School of Hygiene and Tropical Medicine

Melissa Goldstein, JD
George Washington University, Milken Institute School of Public Health

Kathy Hageman, PhD, MPH
CDC, Division of Global HIV, AIDS, TB

Nina Hasen, PhD
Global Health Consultant, ACHIEVE Innovations

Micheal Ighodaro
Prevention Access Campaign, Washington, D.C.

Thoko Kalua, MBBS, MSc
Center for International Health, Education, and Biosecurity (Ciheb), University of Maryland

Jennifer Miller, PhD
Yale School of Medicine

Yogan Pillay, PhD
Department of Health, South Africa

Anton Pozniak, MD
Chelsea and Westminster Hospital, HIV Medicine; London School of Hygiene and Tropical Medicine

Miriam Rabkin, MD, MPH
Columbia University, ICAP Global Health, Systems Strategies

David Ribes, PhD
University of Washington, Dept. of Human Centered Design & Engineering, Data Ecologies Lab

Lisa Singh, PhD
Georgetown University, Massive Data Institute, Dept. of Computer Science

PROJECT TEAM

Charles Holmes, M.D., MPH, Co-PI
Georgetown University, Center for Innovation in Global Health

Maggie Little, Ph.D., Co-PI
Georgetown University, Ethics Lab

Heidi Weimer, J.D., MPH, LL.M, Project Manager
Georgetown University, O'Neill Institute for National & Global Health Law

Dylan Green, MPH, Subject Matter and Technical Expert
Cooper/Smith, Research and Science

Alicia Patterson, Ph.D.
Georgetown University, Ethics Lab

Jonathan Healey
Georgetown University, Ethics Lab

Sydney Luken
Georgetown University, Ethics Lab

INTRODUCTION

Key priorities of HIV programmes include improvements in the ability to identify those most at risk of acquiring HIV, experiencing treatment interruption, or in need of more support to remain on treatment—all in service of providing access to appropriate services. In support of these goals, there is increased interest by HIV researchers and programme innovators in deploying the new capabilities offered by Big Data, including data generated by individuals' use of digital services and devices, such as social media platforms, apps, and mobile phones. In order to leverage the power of Big Data, HIV programmes are using Machine Learning (ML) and Artificial Intelligence (AI) methods to develop models that can better predict which clients or patients might be living with HIV, at risk for treatment interruption, or otherwise underserved. Advances in Natural Language Processing (NLP) models has also led to interest in extracting information about patients' personal life experiences that may be contained as clinician notes in their Electronic Medical Records (EMRs).

While these strategies have strong potential for improving predictive models, they also bring with them critically important ethical questions. These include issues around privacy and justified surveillance, risks to individuals and groups should the data that is harvested and analytically generated become known, and the potential for algorithmic bias. It is thus critical to appreciate the distinct ethical issues posed by these new modes of surveillance and analysis, and consider frameworks for their responsible use.

International guidelines provide important ethical principles relevant to these questions, including guidance on surveillance in public health ethics,[1] the use of AI in health contexts,[2] and data privacy and protection.[3] The purpose of this guidance is to move from broad principles to pragmatically identifying and confronting considerations, complexities, and factors in the context of HIV. Its goals are to 1) help guide researchers and programme innovators in building designs that are ethically, as well as technically, feasible; 2) assist those charged with assessing proposals for funding and implementation to effectively assess the proposed use of Big Data and ML; and 3) identify actions HIV funders and multilateral organizations can take to advance responsible approaches to the use of novel data and ML models in HIV research and programmatic innovation.

This guidance is the product of an 18-member international, interdisciplinary, and intersectoral Working Group, supported by the Bill and Melinda Gates Foundation. The Working Group convened remotely for three workshops. Each workshop extended across three days; with half-day sessions. Members included HIV researchers and programme innovators with experience in novel data and machine learning; technical experts from computer, information, and data sciences; experts in global public health policy and health law; experts in data ethics and ethics of AI; and members from the community of people living with HIV who have been leaders in advocating for their community. Countries represented were Malawi, Mozambique, South Africa, the United Kingdom, the United States, and Zimbabwe. A series of briefings, extensive virtual table-top exercises, break-out groups, and anchoring case studies informed extensive and iterated plenary discussions.

---

[1] World Health Organization, "WHO Guidelines on Ethical Issues in Public Health Surveillance," Geneva: World Health Organization; 2017. License: CC BY-NC-SA 3.0 IGO.

[2] World Health Organization, "Ethics and Governance of Artificial Intelligence for Health: WHO Guidance," Geneva: World Health Organization; 2021. License: CC BY-NC-SA 3.0 IGO

[3] United Nations Development Group, "Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda," resolution 45/95, November 2017, https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda. Digital Medicine Society's Toolkit for Inclusive Digital Health Measurement Product Development: https://datacc.dimesociety.org/development/

KEY ETHICAL CHALLENGES

Proposals to use data from individuals' use of digital services and devices, or highly personal information from EMRs, and deploy ML methods for their analysis, raise distinctive ethical issues, both in general and in the specific context of HIV. The following are especially critical.

Expansion of Digital Surveillance into Private & Communal Spaces

The ongoing collection of personal data from activities on browsers, platforms, apps, and use of mobile phones can impinge on what have become important sites of personal and social life. These raise important privacy concerns, and can have chilling effects on activities important to democratic activities such as political activism. Information included in EMR clinician notes can contain stigmatized information, such as reports of domestic violence and of suicidal ideation, which also raise privacy concerns, and may have chilling effects on what patients share with their health providers.

Gathering and Generating Sensitive Information

Information gathered from individuals' use of digital services and devices, from web pages accessed to patterns of movement, can include highly personal or potentially compromising information beyond sensitive HIV-related issues. The predictive models based on these data, in turn, generate HIV-related risk scores, and attach them to users' individual accounts. Each of these represents an expansion of informational risk to broad populations.

Risks of De-Anonymization

The use of big data and their associated analytics can substantially increase the ability to infer the identity of individuals in anonymized datasets. Identity can often be recovered when multiple datasets are merged together, as is common practice in big data analytics. The ongoing collection of mobile phone location data also carries especially strong concerns about de-anonymization. Because people's patterns of movement across time are unique, the detailed information on times and locations of the mobile phone's use can translate into the ability to identify the person behind the phone's use.

Fewer Surrounding Protections on Access

Personal data sourced from the use of digital platforms and apps is subject to far less regulation than is health data. Access and analysis of novel data also often involves commercial partners, such as telecommunications, app, or platform companies, which have interests that can be at odds with public health interests, including interests in monetizing shared data, or combining the analytic output with their own, often extensive, data on users.

Potential for Civil Rights Violations

Information on citizens' locations can be used to target them; HIV risk scores attributed to accounts can be used to persecute vulnerable populations. These issues are of particular concern in contexts of high political volatility, and contexts in which certain HIV risk-associated behaviors are criminalized.

Algorithmic Bias & Complexity in ML Models

Data used for training and validation can reflect ethnic, socioeconomic, differential access, or other structural disparities. Using that data to guide future decisions can reinforce those disparities. Many ML techniques, in turn, are sufficiently complex that even those who designed them do not fully know how their predictions are made. These forms of complexity make it more difficult to check for bias in the datasets used to train the models, and impede the ability of communities to appeal consequential decisions made on their basis.

TOWARD ETHICALLY RESPONSIBLE BIG DATA & AI IN HIV RESEARCH & PROGRAMMATIC INNOVATION

In navigating the ethical values at stake in the use of novel data and ML methods, research and programmatic innovators, as well as those who serve as funding or adoption gatekeepers, should address the following processes, standards, and expectations.

1. Embed Ethics Into the Concept Development Process

Concept development involves option scanning, early scoping, and preliminary feasibility assessments. In these activities, ethical feasibility should be added to considerations of technical and institutional feasibility. The options considered should be probed for the ethical risks outlined above. A wide breadth of options should be considered, to help identify those that may represent a better overall balance of values, risks, and benefits.

2. Incorporate Robust Community Involvement

Community involvement in consideration of public health programmes - always critical - assumes particular importance when programmes involve expansion into new and less tested areas of digital surveillance. Proper community involvement includes partnership at early stages in the conceptualization and design of a programme; in evaluations of updated information about post-launch efficacy and risks; and in decisions about potential changes, mitigations, or the decision to end a programme. Community involvement advances in user adaptability and an easier uptake of a programme by the community.

3. Values-Centered ML

Some of the most powerfully predictive models can carry the most ethical costs. Proposals for using uninterpretable models should be assessed, not just by the accuracy advantage they may bring, but whether that advantage is worth the added risks of bias and the social opacity. Those developing ML models should measure bias using fairness measures and reduce any model bias that may arise. Researchers and programmatic innovators should not use ML models whose data and analytical methods they cannot audit.

4. Ensure Robust Data Protection & Governance Structures

Given the sensitivity of the data involved and its linkage to individual accounts, it is critical that structures of data protection and governance be especially robust. Appointments and structures of data governance boards should be designed in ways that maximizes independence of the board and its ability to resist potential pressures for access. Design of technical and infrastructural data protection methods should include scenario planning for assessing possible harms across the lifecycle of the project, and include criteria for halting research or programmes when there is emergent risk.

5. Require IRB/ERC Review for Research

Research that involves collecting and generating highly sensitive attributes linked to individual records should be subject to Institutional Review Boards (IRBs)/Ethics Review Committees (ERCs) and require informed consent. When describing the study and potential risks, care should be taken to go beyond description of the immediate data that will be harvested, and include inferences that can and will be drawn from it, risk scores that will be linked to their individual records, and the risk of de-anonymization.

6. Require Ethical Assessment of Programmatic Proposals

Formal ethical assessment of programmatic proposals is critical. Unlike research projects, programmatic implementation involves population-wide harvesting of personal data, usually without consent or opt-out conditions, on an ongoing and often open-ended manner, with real-world decisions being made on the basis of the predictive analytics. Programmatic proposals should be assessed by both funders and governments for issues around privacy rights, informational risks to both individuals and groups; potential bias and degrees of interpretability; potential political or power-based concerns that may increase risks; and whether adequate community involvement was incorporated into the design and development of the programmatic proposal.

7. Ensure Public Disclosure & Accountability

A core tenet of public health ethics is the need for public transparency, justification, and accountability of public health surveillance programmes. Public health authorities and non-governmental organizations engaged in the use of novel data collection and analytic methods must disclose ongoing data harvesting; the use to which it will be put; the results of monitoring and mediation activities; and provide meaningful opportunity for public input, with special attention to those most directly impacted by the programmes.

RECOMMENDATIONS FOR FUNDERS & MULTILATERAL HIV ORGANIZATIONS

The following recommendations outline concrete and immediately actionable recommendations for HIV funders and multilateral organizations to help advance responsible approaches to the use of novel data and ML models in HIV research and programmatic innovation.

1. Develop Concrete Programmes to Support Community Agency

There is an urgent need for research and pilots of models of supporting meaningful community involvement in consideration of these potential projects. Multilateral HIV organizations and funding partners should support the development of programmes and translational tools, co-designed with community representatives, that would help to empower communities to participate in design and decision-making around proposals for these programmes.

2. Increase Expertise in Settings and Organizations where Solutions Are Deployed

Multilateral HIV organizations and funding partners should support briefings and capacity building for Ministry of Health staff on the technical and ethical issues surrounding data collection and analysis efforts, in order to enhance governments' ability to independently assess proposals. Consideration should be given to developing programmes to help in-country researchers and programmatic innovators to become a resource for their countries by increasing their ability to initiate appropriate projects, and their ability to assess projects proposed by others.

3. Convene International Meetings

Major HIV policy institutions and funding partners, such as the World Health Organization, Joint United Nations Programme on HIV and AIDS, The Global Fund to Fight AIDS, Tuberculosis, and Malaria; and the United States President's Emergency Plan For AIDS Relief (PEPFAR), should convene meetings relevant to their remits to bring attention to the ethical issues that surround the use of novel data and ML models. Convenings can include key representatives from the community, implementers, government, and the technology sector to further develop suggestions, recommendations, and normative guidance for reducing ethical risks.

4. Incorporate Ethical Requirements into Funding Proposals

Funders of research and programmatic proposals should require an ethics risk assessment at the full proposal stage. Such assessments should include identification of privacy concerns and context-sensitive risks of informational harms, de-anonymization, and civil rights violations . Funders should expand their review capacities as needed to ensure inclusion of the expertise needed to review these features.

5. Provide Access to Technical & Ethics Advisors

It is important to support the inclusion of ethics and technical advisors in the development and execution of projects. Experts in these domains with experience in these sorts of complex projects are especially valuable. It can be difficult for teams to identify and recruit such individuals on their own. Multilateral organizations and their funding partners should work to develop access routes to such experts. Examples could include establishing a centralized group of relevant advisors to serve as a resource for proposal developers. In addition, the rapidly escalating sophistication of these data harvesting and analytic methods has surpassed the more usual clinical or social science expertise currently populating IRB/ERCs. Training and provision of access to the availability of expert consultants that can help in the review of these protocols.

6. Build Translational Tools For Researchers & Programmatic Innovators

Multilateral HIV organizations and funding partners should develop templates, models, and tools to assist project teams in implementing the suggestions in this guidance, as well as the requirements that funders should issue. Examples could include a model of, or a suggested process for, developing a robust data governance plan, and assessment tools for ethical risks at individual and community levels across the life-cycle of proposed projects.