

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

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.

RECOMMENDATIONS FOR HIV MULTILATERAL ORGANIZATIONS & FUNDERS

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. 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 decision-making around these programmes.

Funders should support development of translational tools that would help to empower communities to participate in decision-making around proposals for novel data analytics programs. Tools should be developed by working groups that include not only technical content experts, but those with expertise in community-centered design, and members of the community.

Funders should consider funding the training of community representatives and advocates to help support the creation of a generation of voices equipped to defend the interests of their communities vis-à-vis the increasing use of data collection and analysis. Funders should also consider leveraging the methods that researchers for HIV vaccines and cures use to inform their workshops and symposia that actively put community advocates in the room with researchers.

Funders should consult with relevant community organizations and advocacy groups to explore what support would be most helpful to promote their ability to influence and assess proposals for the use of novel data and ML models to guide resource distribution decisions. Questions to ask these groups might include what translational tools might be most helpful.

One such example might be a “Smart Questions” tool, which would be co-designed with community advocates, to help translate technical aspects of proposals and their stakes for the community. Such a tool provides key questions for community members to ask of proposals to empower communities to participate in decision-making around the technical issues involved in these programs. Such a tool would also help such groups to confer with and advocate with respect to donors and government agencies. Tools developed for community inclusion in research, such as those outlined in Good Participatory Practice (GPP) Guidelines, may be helpful in building further models for community inclusion in programmatic design and decisions.

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.

Funders should consider 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. Such work might include piloting summer schools that provide education on the technical and ethical issues surrounding the use of Big Data and ML models for up-and-coming in-country researchers.

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. Funders should consider developing training modules for in-country ERCs on the technical and ethical issues surrounding these systems, to increase ERCs' capacity to provide effective oversight of research proposals, including when outside experts should be consulted in the review process.

3. Establish Processes for the Ethical Evaluation of Programmatic Funding Proposals

Funders of programmatic projects should develop and institute processes and procedures for structured ethical evaluation of proposed projects.

In support of this aim, funders should require provision of a formal ethics risk assessment at the full proposal stage. Such assessments should address identification of privacy concerns, including risks of informational harms, de-anonymization, and civil rights violations; and identification of potential bias or problematic uninterpretability of the ML model. They should include identification of political or power-based concerns that increase the potential for misuse of the data or analytic engine, and whether adequate community involvement was incorporated into the design and development of the programmatic proposal.

Proposals should provide clear, accessible descriptions of features relevant to ethical assessment, in language accessible to reviewers from different disciplines of public health, HIV, data science, ethics, and computer science, along with community representatives involved in the review process.

Project proposers should also be required, as a condition of funding, to incorporate ethics/safety planning and implementation activities throughout their projects. Funders should prioritize projects that have a clear ethical risk assessment and decision-making processes outlined throughout their timelines.

Funders should require, and provide resources to enable, effective community inclusion in projects that they fund. Funders should consider requiring that a community expert be involved in the project as a technical advisor from the project's beginning. Such representatives—as with any other experts—ought to be financially compensated for the time they are asked to dedicate to ongoing projects. Dedicated funding for community involvement should be built into the grant mechanism.

Finally, funders should consider giving preferential funding to proposals that incorporate robust and innovative community partnerships across the lifecycle of their projects, including community organizations close to the ground, and strive to make funding available to enable those partnerships.

4. Expand Ethical & Technical Support for Researchers & Programmatic Innovators

The novel data harvesting and analytic methods at issue in this guidance present a combination of technical and ethical risks and issues that many teams are not immediately equipped to handle. Because of this, it is important to support the inclusion of ethics and technical advisors/consultants or team members that project teams may identify for their own proposals.

That said, for many teams, it can be difficult to identify and recruit experts 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 advisors and consultants with the relevant expertise to serve as a resource for proposal developers. Experts in these domains with experience in developing these sorts of complex projects (and not just those who specialize in end assessments) are especially valuable.

Multilateral HIV organizations and funding partners should also support development of templates, models, and tools to assist project teams in implementing the suggestions in this guidance. Examples could include a model of, or a suggested process for, developing a robust data governance plan, and an updated risk assessment tool that is commensurate with the wide reach of AI systems. Such an assessment tool would help diversify the types of risks considered, including by assessing risks at both community and individual levels, risks by time scale (e.g., short, medium, and long term), and risks which are “one off” versus cumulative.

5. 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. Meetings of key representatives from the community, implementers, government, and the technology sector can further develop suggestions, recommendations, and normative guidance for reducing ethical risks.

For instance, the WHO's Department of Global HIV, Hepatitis and Sexually Transmitted Infections Programmes could hold a scoping meeting on the subject of Big Data to put the topic on their radar—and on the radar of all who look to the Department for guidance. UNAIDS could, for instance, gather its country and/or regional teams to provide case study examples of where Big Data is already used in the fight against AIDS and where its resources would be most welcome. UNF's DIAL could use its expertise to present ethical exemplars from the projects they support.

