



DISAGGREGATED DATA CAN REVEAL SYSTEMIC INEQUITIES. IT CAN ALSO REINFORCE THEM — HERE'S HOW.

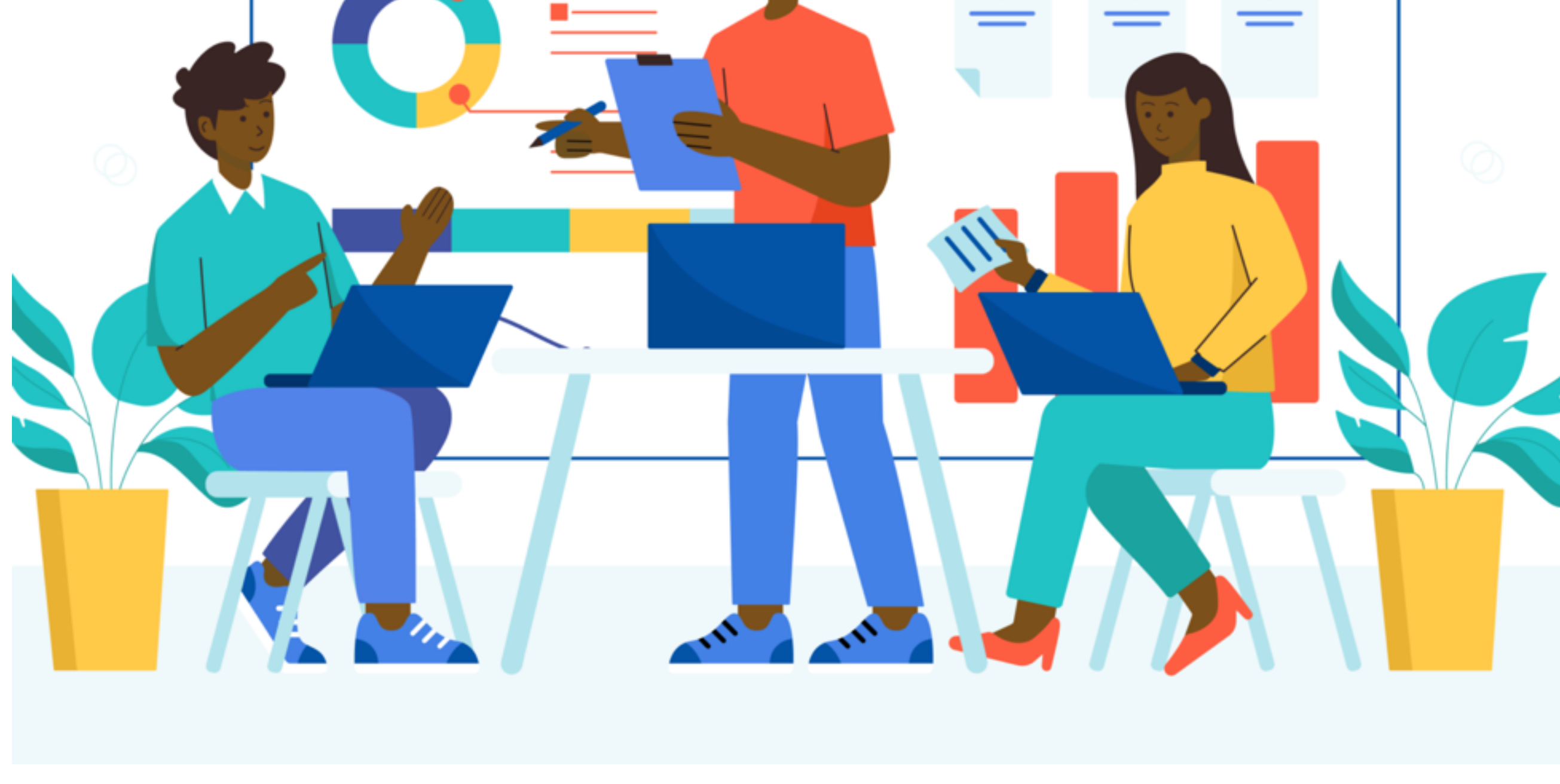
Dec 2, 2021

Authors Brennan Doherty & Avanthika Anand

Experts say disaggregated data collection needs to be done with the consent and guidance of subjects, as well as a genuine commitment to change.

WHY IT MATTERS

Community trust in social impact organizations who do disaggregated data collection can be shattered if it is done badly, especially with subjects who are already overpoliced and surveilled.



Toronto Public Health's (TPH) decision to collect race, ethnicity, income, and housing data from COVID-19 patients to track the virus's spread goes all the way back to the first few months of the pandemic.

The results are unsurprising to anyone with the most basic understanding of structural racism. Around 73 percent of all COVID cases in Toronto, and 74 percent of those admitted to hospital for treatment, are among racialized people. Some of the city's poorest neighbourhoods are among the most ethnically diverse and have the highest COVID-19 case and fatality rates.

Yet Toronto Public Health's decision to collect disaggregated race-based health data was not universally applauded. In June 2020, after TPH published a neighbourhood level COVID map, the Black Public Health Collective condemned the practice in a statement. They pointed out that Black communities and researchers alike are well aware of the inequities they face, and believe race-based data collection isn't just unhelpful — it's harmful.

"There is little evidence that health care or public health systems have used this data to materially change our lives for the better," the collective said in their statement. "Instead, data extracted from Black communities by the state has been used to harm us. We have no reason to believe race-based health data will be anything more than another weapon in the armory of state-implemented anti-Black operations, one that will cause violence far beyond this moment."

Disaggregated data is never neutral. Nor is its interpretation confined to whoever decides to collect or publish it. Sector leaders and privacy experts who spoke to Future of Good say non-profits, charities, grantmakers, and government agencies keen on collecting race-based data should do so in partnership with the communities they're studying. **These communities should have control over not only data collection, but how their data is used.** Organizations also need to understand the privacy risks of collecting disaggregated data, especially on vulnerable populations.

And sector organizations, if they do collect disaggregated data, shouldn't just publish the results in an internal report or review. They should use it to drive change: in program design, advocacy campaigns, or marketing. "For it to be ethical, it fundamentally has to be accompanied by a process that supports the purposes that it's being used for," says Brenda MacPhail, director of the Canadian Civil Liberties Association's privacy, technology, and surveillance program. "That purpose is explicitly to reduce systemic racism, to reduce oppression, to achieve equity and equality."

RESEARCH SUBJECTS MUST BE IN CONTROL

When the Homeless Services Association of B.C. does detailed counts of the province's homeless population, it doesn't just dispatch its own researchers to visit tent encampments and shelters. It sends out a mix of volunteers, staff, and peers to build relationships with homeless people across B.C. and collect data on gender, Indigeneity, and race. While its latest annual survey isn't out yet, past data suggests that B.C.'s homeless population is disproportionately Indigenous: the result of centuries of colonialism and intergenerational trauma.

Stephen D'Souza, executive director of the Homeless Services Association of B.C. says they collaborate with communities they study because it gives the community control over how their information is used and presented. "When you're collecting information from people, those people own that data at the end of the day," he says. "They should have the right to influence the data that comes from it."

The Black Public Health Collective concurs. "Our own communities are much better suited and trusted to collect data for the purposes of ending health disparities," they wrote in their June 2020 statement. Creating processes that allow research subjects to exercise self-determination as equals, rather than the traditional 'objective' distance between subject and researcher, would allow for a degree of accountability.

When the Homeless Services Association of B.C. is collecting data, D'Souza says, working with subjects allows them to tell researchers why they're noticing certain trends in data. "That's the context that needs to come with the data, not us just making assumptions as those outside of that data," he explains. "Interpretation and collection of that data needs to come from the community that you've collected that data from."

DE-ANONYMIZED DATA IS A RISK

Disaggregated data is touted as a relatively risk-free way to balance an understanding of how race, ethnicity, or other personal demographic details affect quality of life with a need for privacy. And rightly so — while understanding racial inequity is important, there is a risk that disaggregated data can be used to identify subjects at risk of violence because of their identity. The Black Public Health Collective points to the use of an individual's COVID-19 data by the Toronto Police Service and the Toronto Transit Commission to track them.

As MacPhail explains, the process of de-identifying data to strip out the personal identifiers of individual subjects isn't perfect. "That's a complex process," MacPhail explains. "There is serious debate and, I think, an emergent understanding in the technical world that it is literally impossible to 100 percent successfully de-identify data."

Stripping away obvious identifiers like names, ages, and birthdays might be easy, but proxy identifiers like postal codes (relevant for a lot of population health data) can still tell a lot about a subject and could, theoretically, be used to identify them — or, at the very least, single out their neighbourhood. **"Existing COVID-19 data has already been used to assign greater police presence in COVID hotspots," said the Black Public Health Collective's June 2020 statement. "These related narratives have pathologized Black communities as the 'sick' and reinforced harmful narratives about us."**

MacPhail says the risk is especially high for non-profits and charities who might have excellent intentions and a great data collection process, but due to a lack of resources, may not have the most rigorous 'de-identification' process to strip away personal identifiers. "You have to do due diligence when you're choosing what kind of process you're going to use to de-identify the data," MacPhail says. Good processes will be transparent, verifiable, and testable for rigour. But even the most perfect disaggregated data collection process will only go so far.

"De-identification alone is necessary", MacPhail explains, "but it's insufficient when we're dealing with this kind of data about communities who have historically suffered from systemic racism, from oppression, and from appropriation of information about them in ways that have historically done them harm."

DISAGGREGATED DATA MUST PROMPT POLICY CHANGE

Simply collecting disaggregated data on a marginalized population isn't enough to guarantee racial equity. Case in point: statistics collected internally by the Toronto Police Service for decades through the practice of carding showed just how racist the force was (and still is) towards Black and brown Torontonians. The collection of this data didn't change anything. Only after multiple investigations by the Toronto Star and other news outlets used this data to shame the force did its leadership publicly commit to change.

Critics of the recent wave of calls for more race-based disaggregated data collection point out the frequent disconnect between data on racial disparity and improvement to the living conditions of racialized communities through policymaking. In an essay on the website of the Royal Society of Canada, Rinaldo Walcott, director of the University of Toronto's Women and Gender Studies Institute, points out that concrete policies to address social inequities rarely follow the heels of race-based data collection. "It is a half-truth to link data and in particular race-based data collection with good policy-making," Walcott wrote.

Dana Riley, program lead on the population health team at the Canadian Institute for Health Information, says her work on disaggregated data collection suggests race-based and Indigenous identity data is very necessary, but it needs to be collected for a purpose. "We want to ensure that if we're going to collect this type of data that it's used to improve the lives and the healthcare outcomes of racialized and Indigenous groups," she says. Doing so requires social service organizations to not only put their data where their mouth is — by adapting their services in response to data — **but also be transparent and accountable to their subjects on how their data will be used.**

At the Homeless Services Association of B.C., D'Souza says their data on the province's homeless population is used to help service providers understand who isn't using their services and how they can be brought into the fold. It helps governments tweak their social programs for homeless people using their services — for example, those leaving a hospital after treatment. "We have used disaggregated data from both ends," D'Souza says, "both trying to understand who's in the community and better support government and service providers around who's not in their spaces — and the work that needs to be done."

WHAT'S IT ALL FOR

Data collection, privacy, and the autonomy of subjects are contentious topics in the research world. Good disaggregated data collection coupled with rigorous ethical standards, community participation, and thoughtful presentation can highlight severe racial disparities. Or, it can be fodder for racist caricatures of a community.

As non-profits, charities, grantmakers, and government agencies increasingly see race and ethnicity as a relevant factor in solving social inequities, they need to be extremely careful about how they pursue disaggregated data collection. Respecting the autonomy and agency of subjects is among the most important values a researcher can have, as well as a robust process for de-identifying data.

But above all, Walcott says, researchers need to understand that disaggregated data is simply a means to an end — be that for social purpose organizations or communities. "All data can do is inform policy-making, if anything at all," he wrote in his essay. "Policy-making, afterall, is ultimately about political decisions."



Brennan Doherty is the Feature and News Writer for Future of Good.



Avanthika Anand is the News and Features Writer at Future of Good.

READ THIS NEXT



TECHNOLOGY
CANADA LAGS BEHIND THE US ON DISAGGREGATED DATA — BUT THESE CANADIAN ORGANIZATIONS ARE CATCHING UP. HERE'S WHAT THEY'VE LEARNED.

By Avanthika Anand — Dec 1, 2021



DATA
THREE WAYS COVID RESPONSE COULD HAVE BEEN BETTER WITH (MORE) DISAGGREGATED DATA

By Brennan Doherty — Nov 22, 2021