AI to Save Lives on City Roads

GUADALUPE BEDOYA, ARIANNA LEGOVINI, AND SVETA MILUSHEVA

Common Perceptions

Road traffic crashes (RTCs) devastate the lives of victims and their families, while causing economic destruction in terms of property and lost earnings, especially for the poor. With 1.35 million deaths per year (World Health Organization 2018), RTCs are the primary cause of death among young people (World Health Organization et al. 2010), and the eighth leading cause of death globally (World Health Organization 2018). Each year, they cause an estimated 20–50 million nonfatal injuries globally, imposing heavy burdens on health systems (World Health Organization et al. 2010). The overall annual cost to countries is estimated between 1 and 2 percent of GDP.

Policy action should focus on Africa, where road traffic deaths are the highest (averaging 26.6 deaths per 100,000 people), and notably in contexts like Kenya, where rates rank among the highest in the world (World Health Organization 2018). These fatalities disproportionately affect the poorest inhabitants who rely on walking as their main commuting method. Pedestrian deaths in Nairobi represent 71 percent of all RTC deaths reported in police crash records. The Kenyan government's ability to develop effective road safety interventions has been historically constrained by poor quality data, limited analytics, and a lack of policy experimentation.

Questions We Should Be Asking

The Stockholm Declaration by the Third Global Ministerial Conference on Road Safety reiterated the call for increased country investments in road safety—from legislation and regulation, safe urban and transport design, and safe modes of transport and vehicles, to modern technologies for crash prevention, trauma care, and urban management. Limited resources, however, make it unlikely that countries will be able to implement all of these recommendations across their entire road network. Instead, countries should invest where it matters most. This requires knowing where and when crashes happen, so that resources can be targeted to risky locations and times.

However, data gaps remain a fundamental limitation. Only 17 percent of road traffic deaths are estimated to be reported by low-income countries (World Health Organization 2018). Furthermore, available reports are often not digitized and are therefore difficult to analyze. In Kenya, the World Health Organization estimates the number of RTC fatalities to be 4.5 times higher than what is officially reported (World Health Organization 2018). What other data sources can decision-makers use to fill such data gaps? Can the large quantities of privately held data generated from smartphones and social media help geolocate RTCs? DIME's ieConnect for Impact team partnered with the Kenyan government to answer these questions; develop evidence-based, actionable policy insights; and improve the country's road safety and transport policy.

Challenging Perceptions

The partnership set out to produce analytics that could help identify adverse events, flag risk factors, reduce reaction time, and improve planning. The strategy adopted was to: invest in innovative ways to collect high-frequency, real-time data from multiple sources; understand

local constraints to road safety; and experiment with behavioral, infrastructure, and institutional approaches to improve road safety.

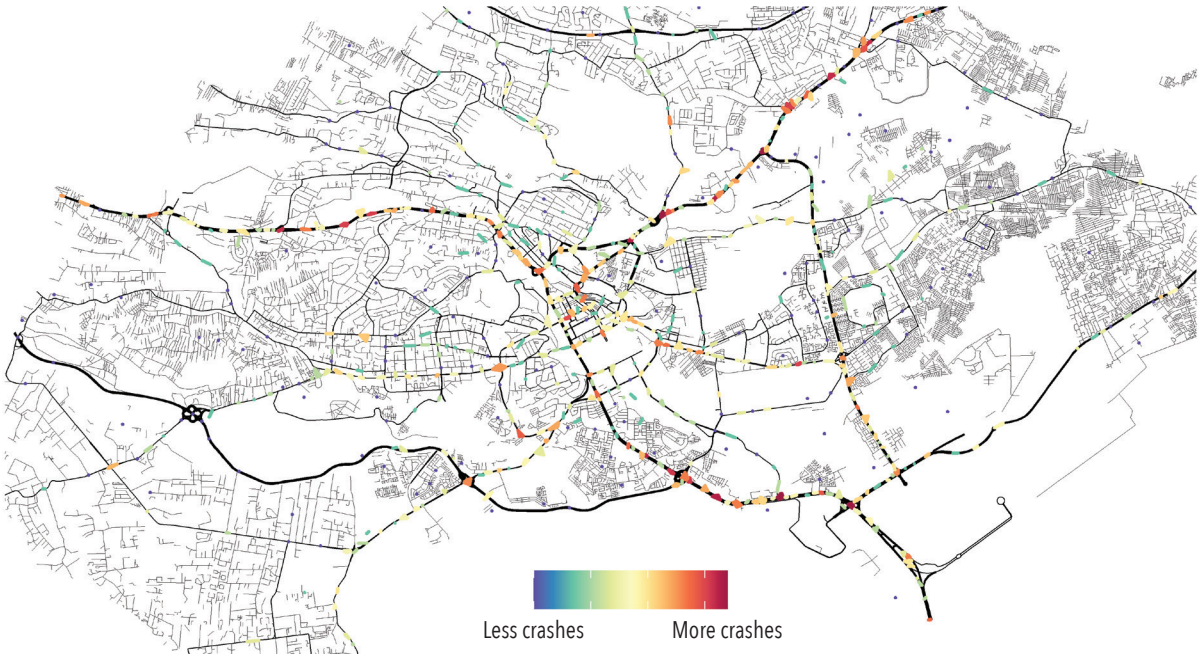
To meet the first objective, DIME combined thousands of administrative records (paper-based crash reports generated by the Kenyan National Police Service and digitized by the DIME team) with crowdsourced data from Twitter (see map 4.1). DIME researchers developed machine learning algorithms using natural language processing to geolocate the RTCs reported on Twitter. The data was validated in real-time by a motorbike dispatch service, which confirmed the accuracy of the geoparsed tweets in 92 percent of cases in the validation sample. Merging police and crowdsourced data showed at least 50 percent more crashes reported by bystanders than in police crash reports.

The second step was to use this new information to better understand how and why accidents happen. Although crashes occurred in 1,400 locations across the city, half of them were concentrated in clusters representing less than one percent of the road network (see map 4.2). Fifty-three percent of deaths and 50 percent of injuries occurred early in the morning (5–8 a.m.) and later in the day (5–11 p.m), highlighting the potential roles of low visibility, low alertness, and alcohol. Combing through the data also revealed that 35 percent of deaths occur within twenty meters of bus stops. The particularly hazardous locations where crashes frequently occur are referred to as blackspots. A diagnosis of their typology and associated risk factors was established by analyzing physical attributes such as the presence of sidewalks, streetlights, safety barriers, and pedestrian crossings. The team collected these in 200 high-risk locations using a survey that adapted the principles from iRAP (the lead organization for road assessment programs worldwide) and combined them with additional data collection on behavioral risk factors.

Map 4.1 Nairobi Crash Map (individual events)

Note: The map displays crash reports in Nairobi, Kenya, obtained from digitized police records (in purple) and crowdsourced Twitter data (in green).

Map 4.2 Nairobi Crash Map (clustered)



Note: The map displays clustered reports of road traffic crashes in Nairobi, Kenya, obtained from digitized police records and crowdsourced Twitter data. 1% of the road network accounts for over 50% of all crashes, deaths and injuries.

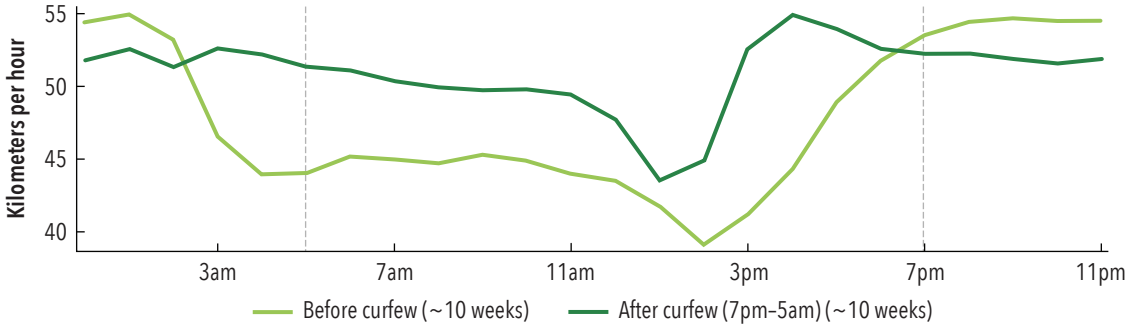
Additional data from a variety of platforms were integrated into the crash reports to produce an even richer picture. For example, Uber and Waze data provided information on average speeds on different road segments and on obstacles such as potholes. Weather data was used to understand driving conditions, and Google Maps data was used to help researchers understand land usage such as pedestrian traffic near educational and health facilities. This combination of high-frequency, spatially integrated data systems led to numerous actionable, real-time insights. For example, when the city of Nairobi implemented curfew measures to limit the spread of COVID-19, the change in average speed as people hurried home became immediately apparent (see figure 4.2). Because speed is a key risk factor for RTCs, these changes carry important implications for agencies looking to prevent crashes.

Policy Implications

The collaboration between DIME and the Kenyan authorities demonstrates that government policy can be effectively bolstered by building ecosystems that integrate machine learning algorithms with survey, administrative, and new data sources. In the case of Kenya, crowdsourced and private mobility data were merged with official records to build the first georeferenced, multi-year crash dataset for the city of Nairobi. The tools and instruments developed in this pilot can be deployed across Africa to facilitate appropriate planning and optimize resource allocation to improve road safety.

Analyzing such data shows how critical it is to clearly identify problems and prioritize policy interventions around them. In developed countries, safety strategies aimed at risky locations reduced the

Figure 4.2 Nairobi Crash Map (clustered)



Note: The graph shows the impact of implementing a COVID curfew on the average speed of commuters in Nairobi, Kenya. Road users significantly increased their average speed as they hurried home to avoid checkpoints before nightfall. The lower number of road users also contributed to higher speed averages.

concentration of crashes: most fatal and serious crashes now occur at locations on roads with no other injury crashes reported in the previous five years. Therefore, in countries where crashes are concentrated, which is likely the case for the deadliest countries in terms of RTCs, blackspots should be the focus of road-safety analysis and policy.

Such a strategy avoids dissipating scarce fiscal resources and instead uses them where policy action can most effectively reduce mortality. Targeting 1 percent of the road network in Nairobi where severe crashes have occurred most frequently and focusing enforcement efforts in the most dangerous hours of the day could potentially halve the number of deaths. Moving from addressing road safety on the entire road network to focusing on a relatively small portion reduces an intractable problem to a more manageable one, and can help cities reach their 2030 SDG target.

The Kenya Urban Roads Authority, the agency responsible for city infrastructure, used the data to inform the selection of 102 high-risk locations to install speed cameras and upgrade infrastructure.

The Kenyan example demonstrates how proper documentation and data extraction can create a roadmap to build innovative data systems in data-scarce settings to monitor and manage road safety policy, and support policy making. The wider significance of this work points to the possibility of harnessing time-stamped geo-located data and statistics on different events that are either recorded by the government or reported on social media. These improved tools could help geolocate victims during a natural disaster, locate crime scenes, or pinpoint areas needing immediate attention. Identifying the time and location of events can thus contribute to automating and accelerating policy responses across a wide set of contexts and issues that affect people's lives.

This case study is based on an impact evaluation conducted within DIME's Infrastructure and Climate Change research program. See: Milusheva, Sveta, Robert Marty,* Guadalupe Bedoya,* Elizabeth Resor, Sarah Williams, and Arianna Legovini.* 2021. "Applying Machine Learning and Geolocation Techniques to Social Media Data (Twitter) to Develop a Resource for Urban Planning." PLOS One 16 (2): e0244317.*

REFERENCES

Milusheva, Sveta,* Robert Marty,* Guadalupe Bedoya,* Elizabeth Resor, Sarah Williams, and Arianna Legovini.* 2021. "Applying Machine Learning and Geolocation Techniques to Social Media Data (Twitter) to Develop a Resource for Urban Planning." *PLOS One* 16 (2): e0244317. https://doi.org/10.1371/journal.pone.0244317.

World Health Organization. 2018. *Global Status Report on Road Safety 2018*. World Health Organization. https://www.who.int/publications/i/item/9789241565684.

World Health Organization, FIA Foundation for the Automobile and Society, Global Road Safety Partnership, and the World Bank. 2010. *Data Systems: A Road Safety Manual for Decision-Makers and Practitioners*. World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/44256/9789241598965_eng.pdf?sequence=1andisAllowed=y.

ACKNOWLEDGMENTS

The research team comprises Guadalupe Bedoya,* Arianna Legovini,* Robert Marty,* Sveta Milusheva,* and Sarah Williams. This work is a collaboration between the World Bank Transport Global Practice, the National Transport and Safety Authority, the National Police Service, and other Kenyan government agencies. Amy Dolinger, Meyhar Mohammed, and Robert Tenorio provided research assistance throughout the project. Elizabeth Resor developed the first geocoded crash map and provided field support. Funding was provided by the DIME Impact Evaluation to Development Impact ieConnect program, which has been funded with UK aid from the UK government, the Transport Global Practice at the World Bank and the World Bank's Knowledge for Change Program.

**Development Impact Evaluation (DIME), World Bank.*