# **Potential Projects (Please add to it)**

1. **Harnessing population sources:** Develop a tool (Excel, R, Shiny, etc.) that would allow ICPI analyst and SI staff on field teams to easily toggle between Census, WorldPop, and LandScan subnational (SNU1, SNU2, etc.) population denominator data. This tool would aid ICPI staff in performing sensitivity analyses in areas where existing population denominator data is suspect. We developed a similar tool for CDC-Ethiopia that allowed them to toggle between Census, WorldPop and LandScan population denominator data in the calculation of PLHIV at SNU2 and SNU3. This proved especially useful in areas like Addis Ababa where existing Census data likely did not reflect the current reality.
2. **Clustering sites by simulating patient flow:** Apply treatment gap model(s) to additional countries to determine whether the models are truly able to pin point sites that are struggling to link newly diagnosed HIV+ persons to treatment. This project would involve running the models on PEPFAR data and working with a select set of countries to investigate the accuracy of the results. The impact is potentially huge as it would allow us to identify (and remedy) sites/partners/locations that are struggling to link newly diagnosed HIV+ persons to treatment. This project hinges upon the quality of the HIV testing data and will require working with a country team to label the data to add in the site attributes that define the expected referral patterns and the travel times between sites.
3. **Aberration detection algorithms into Genie:** Integrate aberration detection algorithms into Genie to assist with identifying aberrations/data entry errors. This tool would potentially serve two purposes: (1) improve data quality and (2) serve as an early warning system. This tool can be potentially used for DRT export from the data genie and identify 1-2 sites that are planning on creating their own DRTs on monthly/weekly data.
4. **Bayesian clustering:** In many countries, partners are collecting weekly, even daily data on testing and treatment numbers for individual sites. This wealth of data provides an opportunity to run machine learning models at site levels. One of many ways to cluster sites is to group sites based on self-learning Bayesian clustering models. These micro level clustering models could pick up signals based on geo-location of sites and could be further used to optimize the yield at district/zonal level.