Supplementary Appendix

Geospatial Statistical Analysis

In order to evaluate trade-offs between efficiency and geographic equity in access to health services, we sought to both describe geographic correlations between measures across different countries, and identify geographic clusters of efficient and inefficient allocation of health facilities within countries. Doing so required the use of spatial statistical procedures that are commonly used in fields of geography, demography, and ecology [1–3]. A brief description of these procedures is provided below, along with justification for using these methods to answer our research question.

Testing for global spatial autocorrelation between district-level population density and travel time

We chose to consider spatial autocorrelation in our statistical analysis because unmeasured spatial processes, such as social and political factors, are likely to influence health service distribution and access [4,5]. Spatial autocorrelation refers to the extent to which values of a variable in geographic space vary based on proximity between neighboring units [3]. Positive spatial autocorrelation indicates that neighboring values are more likely to be similarly high or similarly low (high surrounded by high, low surrounded by low). Negative spatial autocorrelation indicates that neighboring values are more likely to exhibit an inverse correlation (high surrounded by low, low surrounded by high). These ideas can be extended to bivariate settings, in which the value of variable x covaries with values of another variable y based on proximity.

Hypothesis tests for geospatial statistics compare observed values in geographic space to those which would have been observed under the null hypothesis of spatial randomness (no spatial ordering to values in space) [3,6]. Implementing these spatial analysis procedures requires neighboring units to be defined using a weights matrix. This weights matrix is often denoted using the notation w_{ij} with equal rows i and columns j, where both i and j are equal to n, the total number of districts in the analysis. Each cell in w_{ij} contains a binary (1/0) element indicating if district j is or is not a neighbor of district i. In our study, we defined neighbors using the "queen" contiguity matrix, in which neighboring districts in any geographic direction would be considered neighbors because we assumed that spatial processes that drive population density and travel times would be expected to be shared across all neighboring regions [6]. Districts without neighbors were excluded from the analysis because they would not contribute to the weights matrix or spatial statistical calculations. Permutation tests were used to determine statistical significance, and we specified 999 permutations [1,2].

We applied the Global Moran's I statistic (Equation 1) to test for global spatial autocorrelation:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Equation 1: Global Moran's I Statistic

Where n is the total number of districts, w_{ij} is the neighborhood weights matrix, x_i is the value of population density for district i, $\bar{x}$ is the global average district-level population density, y_j is the value of travel time in neighboring district j, and $\bar{y}$ is the global average district-level travel time.

Identifying local clusters of positive and negative spatial autocorrelation between district-level population density and travel time

While the global Moran's I test provides evidence of spatial autocorrelation (whether or not the magnitude of population density and travel covary in space in the total study region), it does not allow investigation of the four patterns of correlation between district-level population density and travel time that we were interested in investigating, and that drive the global association. For policy recommendations and planning, local officials must identify where clusters of sub-optimal allocation (high population density, long travel time; low population density, short travel time) are located. Note that these are examples of negative spatial autocorrelation (values of population

density are inversely correlated with values of neighboring travel time). Efficient clusters (high population density, short travel time; low population density, long travel time) are examples are positive spatial autocorrelation.

The Local Indicator of Spatial Autocorrelation (LISA) statistic can be used to identify geographic clusters (Equation 2):

$$I_i = \frac{(x_i - \bar{x})}{\sum_i \frac{(x_i - \bar{x})^2}{n}} \sum_j w_{ij} (y_j - \bar{y})$$

Formula 2: Local Moran's I Statistic

All notation is the same as previously specified. The major difference between equation 1 and 2 is that in equation 2, we obtain a statistic for each district *i*. Heuristically, this method can be interpreted as a regression where district-level population density is the independent variable, and district-level travel time is a spatially smoothed estimate of travel time in its neighboring areas [2]. This method can be used to identify statistically significant clusters in efficient and inefficient quadrants within the Geo-PSA conceptual framework that can be further investigated to inform policy. Because each district *i* through *n* generates its own test statistic, we applied the Benjamini Hochberg False Discovery Rate correction for multiple testing in our primary analysis [7]. As a sensitivity analysis for exploration, we included analyses using the less stringent traditional alpha=0.05 cut-off and have included those results in Supplementary Figures S1 and S2.

Interpretation of geospatial analysis and relevance to evaluating trade-offs between equity and efficiency

In summary, these geospatial statistical procedures made two important contributions to our evaluations of trade-offs between equity and efficiency in the spatial allocation of health services with respect to populations in need. First, tests for global spatial autocorrelation revealed whether the magnitude of district-level population density and travel time tracked together in space. A priori, we hypothesized there would be lower spatial autocorrelation when evaluating primary care services compared to cancer services. Investments in equitable geographic coverage of services could limit the variability in district-level travel times for the whole country, leading to lower spatial autocorrelation between population density and travel time. However, for cancer referral centers, we expected to observe stronger negative spatial autocorrelation because services are more likely to be found in highly populated areas, where shorter travel times would be expected to be found. Regions far from urban centers would be sparsely populated and exhibit longer average travel times to reach cancer care. Second, identifying local clusters of positive (inefficient) or negative (efficient) would allow policymakers to direct resources to those areas to address these imbalances – information to drive action. LISA clusters indicate the specific spatial relationships (efficient or inefficient) between population density and travel time that drive the global correlations, allowing for more rigorous investigation of causes for those imbalances.

Data Sharing Appendix. Analytic Workflow and Data Repository.

In order to enable policymakers and analysts to implement the Geographic-Population Services Access (Geo-PSA) analytic approach, we include additional information regarding specific software and data requirements, as well as a link to data elements and R code used to generate the maps and scatter plots presented in this article.

Analytic Workflow

The analysis presented here includes five main steps:

1. **Preparation of geographic data inputs (administrative boundaries, land cover, elevation, road networks, population density, facility locations).** Sources for these data elements have been provided in Appendix Table 2 below. Particular care should be taken to ensure that all raster data sources use the same projection and resolution. In addition, assumptions regarding travel time and mode of transit for different land cover classes should be informed by local knowledge provided by patients directly, or health care providers with experience working with patients.

2. **Estimating travel time using Access Mod 5**. Users should upload all relevant geographic datasets to Access Mod 5 and perform the geographic accessibility analysis tool. Tutorial videos and documentation are available on their website [8].

3. **Calculating zonal statistics for the average population density and travel time within the administrative boundary of interest**. This can be done directly in Access Mod 5 using the zonal statistics tool, or in ArcMap or QGIS using the zonal statistics as table command.

4. **Estimating the bivariate local indicator of spatial autocorrelation (LISA) and creating indicator variables corresponding to each cluster**. (1: high population density/short travel time, 2: high population density, long travel time, 3: low population density, short travel time, 4: low population density, long travel time). We used the publicly available GeoDa software package developed by Dr. Luc Anselin and currently supported by the University of Chicago. The user can specify application of the Benjamini-Hochberg False Discovery Rate correction for multiple testing when performing the bivariate LISA analysis. The package can be downloaded at their website [9].

5. **Generating summary statistics, maps and scatter plots**. This can be done using any statistical software package of choice. We have chosen to use R because of its superior data visualization capabilities, reproducibility, and ability to work with both geospatial and aspatial data.

Data Repository

We have posted datasets and code used to generate the key results in the main manuscript. Specifically, we have created a GitHub repository containing the travel time raster datasets and the R code used to generate the bivariate LISA scatter plots and maps evaluating associations between population density and travel time to health centers and hospitals. [10]

Appendix Table 1. Using the Geographic-Population Services Access conceptual framework to guide policy response

Quadrant	County Classification	Policy Response
Lower right	High Population Density, Short Travel Time	Efficient, centralized services that serve many people
Lower left	Low Population Density, Short Travel Time	Inefficient, placing resources in poorly populated areas will drive up costs of tests. Consider shifting resources to more populated areas.
Upper left	Low Population Density, Long Travel Time	Efficient but inequitable, policymakers may consider introducing point-of-care testing, telepathology, outreach interventions to facilitate referrals to reduce excessive travel times
Upper right	High Population Density, Long Travel Time	Inefficient, too few resources for high population. If resources are available, centralized health services should be placed in these areas

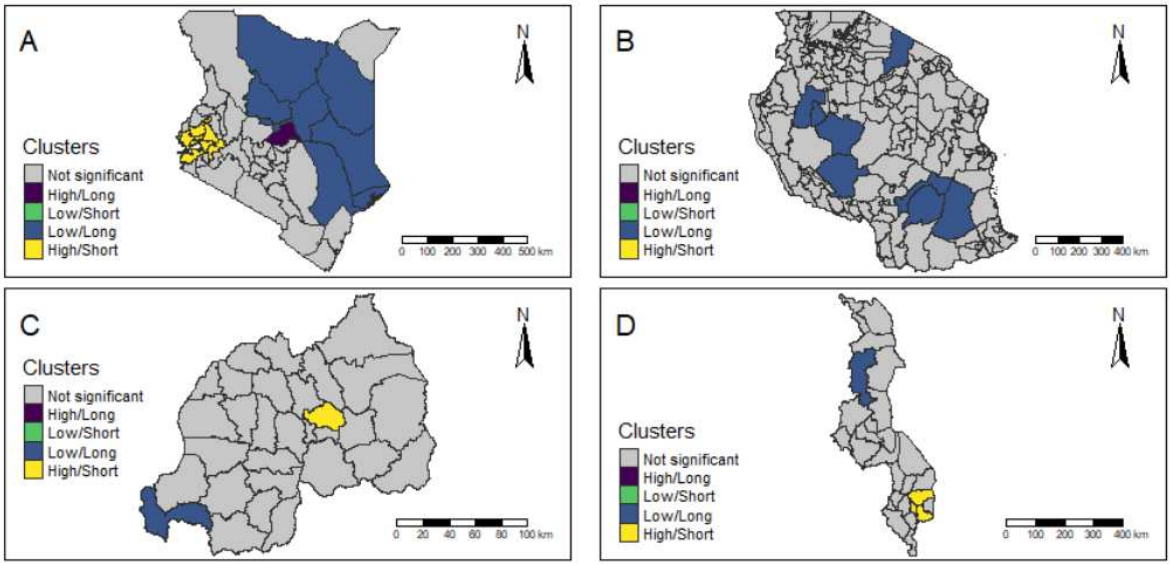
Iyer HS, *et al. BMJ Global Health* 2020; 5:e003493. doi: 10.1136/bmjgh-2020-003493

Appendix Table 2. Geospatial data required for evaluating spatial relationships between population density and travel time

Database name	Description	Time	Resolution	Specific analytic input	Source/citation
MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid (MCD12Q1 v006)	Satellite image database that provides global land cover types at annual intervals from 2001 to 2018.	2014	500m	Land cover (forest, shrubland, wetland, savannah, croplands, urban, barren, water, permanent snow)	USGS Friedl & Sulla-Menashe 2019. Available through Google Earth Engine.[11]
Columbia SEDAC gROADSv1	Global database of road networks harmonized across countries.	1980-2010		Roads	Center for International Earth Science Information Network (CIESIN)/Columbia University, and Information Technology Outreach Services (ITOS)/University of Georgia. 2013. Global Roads Open Access Data Set, Version 1 (gROADSv1). [12]
World Pop Project	Modeled high-resolution population estimates available from 2000 to 2019	2015	100m	Population per 100m^2	World Pop Project Population Counts [13,14]
GMTED 2010: Global Multi-resolution Terrain Elevation Data	Elevation measured using multiple sources, primarily NASA digital terrain elevation data	2010	225m	Elevation in 2010	Global Multi-resolution Terrain Elevation Data 2010 courtesy of the U.S. Geological Survey [15]
ESRI hydrolines and hydropolys	River and Lake boundaries			Line files for rivers and lake boundaries for each country	ESRI ArcGIS catalog [16]
Health facility locations (e.g., all facilities in sub-Saharan Africa)	Listing of government-run health facilities in sub-Saharan Africa	2019		Provides details about level of health system and latitude and longitude coordinates	Maina et al. 2019 [17]

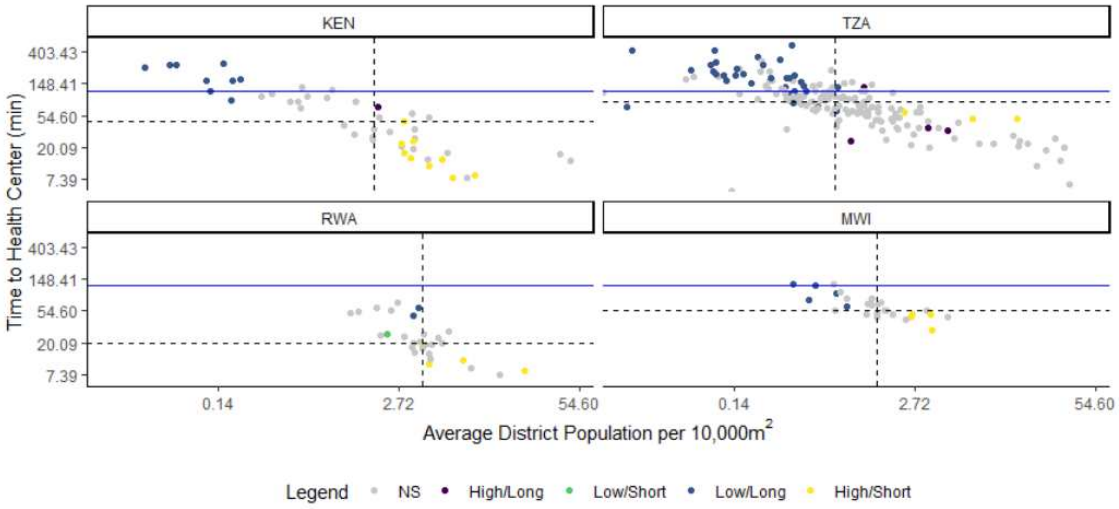
Appendix Table 3. Travel speed and modes of transport for different land cover types specified in the Access Mod 5 algorithm

Label	Speed (kilometers/hour)	Mode
Evergreen Needle Leaf Forests	2	Walking
Evergreen Broad Leaf Forests	2	Walking
Deciduous Needle Leaf Forests	2	Walking
Deciduous Broad Leaf Forests	2	Walking
Mixed Forests	2	Walking
Closed Shrublands	4	Walking
Open Shrublands	4	Walking
Woody Savannas	2	Walking
Savannas	6	Walking
Grasslands	6	Walking
Permanent Wetlands	2	Walking
Croplands	6	Walking
Urban and Built up Lands	15	Bicycle
Croplands Natural Vegetation	6	Walking
Permanent snow and ice	0	Walking
Barren	6	Walking
Water bodies	0	Nothing
Unspecified	20	Motorized
Highway	100	Motorized
Primary	100	Motorized
Local/Urban (Malawi)	40	Motorized
Secondary	60	Motorized
Tertiary	40	Motorized
Trail	20	Motorized



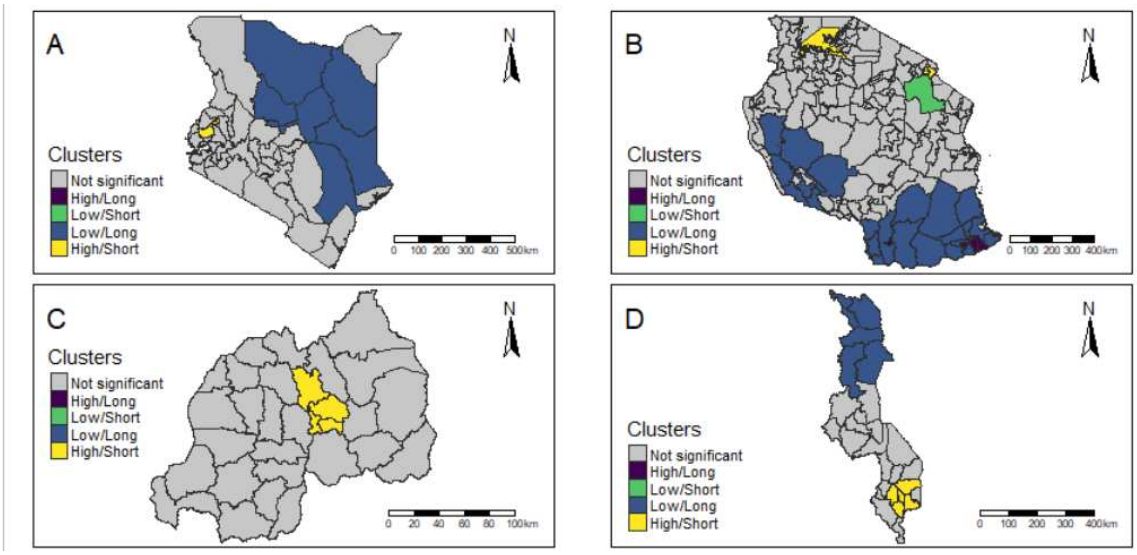
Appendix Figure 1. Maps of bivariate Local Indicator of Spatial Autocorrelation (LISA) plots depicting locations of clusters of four different categories of association between district-level population density and travel time to nearest health center.

Note: The Benjamini-Hochberg False Discovery Rate was applied to correct for multiple testing of clusters. Panel A: Kenya, Panel B: Tanzania, Panel C: Rwanda, Panel D: Malawi. Colors correspond to districts with statistically significant clusters of (1): High population density/long travel time, (2): Low population density/short travel time, (3): Low population density/long travel time, and (4): High population density/short travel time.



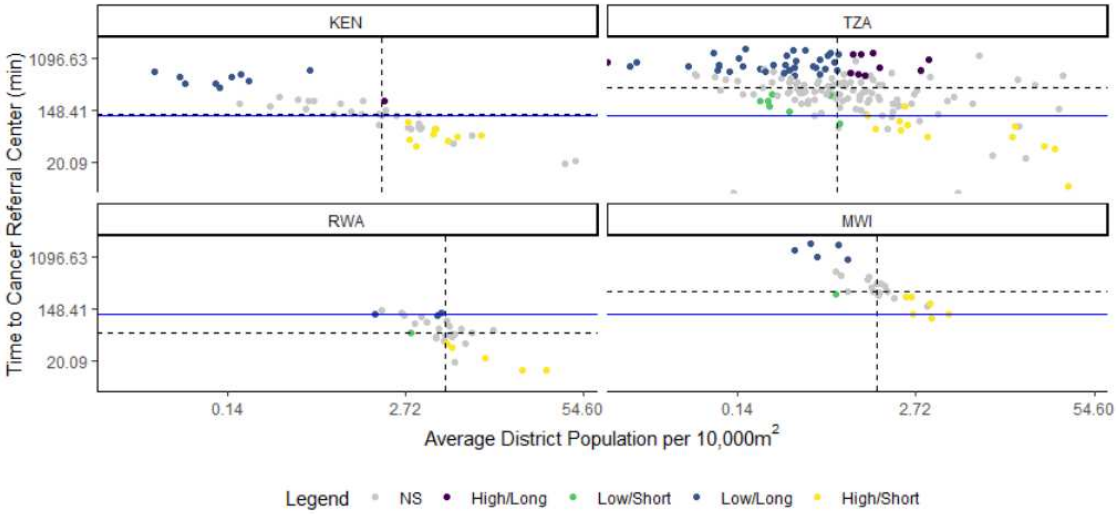
Appendix Figure 2. Pearson correlation between district-level travel time to the nearest primary care health center and population per 10,000m^2 in Kenya, Tanzania, Rwanda, and Malawi.

Note: Abbreviation: NS=Non-significant cluster. Colors correspond to districts with statistically significant bivariate Local Indicator of Spatial Autocorrelation clusters of (1): High population density/long travel time, (2): Low population density/short travel time, (3): Low population density/long travel time, and (4): High population density/short travel time. Significance tests for district clusters were conducted using 999 permutation tests with an alpha=0.05. Blue horizontal line denotes 120 minute travel time. Dotted lines intersect at the median travel time and population density for each country.



Appendix Figure 3. Maps of bivariate Local Indicator of Spatial Autocorrelation (LISA) plots depicting locations of clusters of four different categories of association between district-level population density and travel time to nearest cancer center.

Note: The Benjamini-Hochberg False Discovery Rate was applied to correct for multiple testing of clusters. Panel A: Kenya, Panel B: Tanzania, Panel C: Rwanda, Panel D: Malawi. Colors correspond to districts with statistically significant clusters of (1): High population density/long travel time, (2): Low population density/short travel time, (3): Low population density/long travel time, and (4): High population density/short travel time.



Appendix Figure 4. Pearson correlation between district-level travel time to nearest cancer center and population per 10,000m^2 in Kenya (KEN), Tanzania (TZA), Rwanda (RWA), and Malawi (MWI).

Note: Abbreviation: NS=Non-significant cluster. Colors correspond to districts with statistically significant bivariate Local Indicator of Spatial Autocorrelation clusters of (1): High population density/long travel time, (2): Low population density/short travel time, (3): Low population density/long travel time, and (4): High population density/short travel time. Significance tests for district clusters were conducted using 999 permutation tests with an alpha=0.05. Blue horizontal line denotes 120 minute travel time. Dotted lines intersect at the median travel time and population density for each country.

Reference

1. Anselin L. Local Indicators of Spatial Association—LISA. Geographical Analysis. 1995;27(2):93–115.
2. Anselin L. Exploring Spatial Data with GeoDa: A Workbook. Center for Spatially Integrated Social Science: University of Illinois, Urbana-Champaign; 2005.
3. Cliff A, Ord J. Spatial processes: models and applications. London: Pion; 1981. 266 p.
4. Kim JY, Farmer P, Porter ME. Redefining global health-care delivery. The Lancet. 2013 Sep 21;382(9897):1060–9.
5. Marmot M. Achieving health equity: from root causes to fair outcomes. The Lancet. 2007 Sep 29;370(9593):1153–63.
6. Goodchild MF. SPATIAL AUTOCORRELATION. Concepts and Techniques in Modern Geography. 1986;47.
7. Castro MC de, Singer BH. Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Tests in Local Statistics of Spatial Association. Geographical Analysis. 2006;38(2):180–208.
8. AccessMod 5 | Modelling physical accessibility to health care [Internet]. accessmod. [cited 2020 Mar 4]. Available from: https://www.accessmod.org
9. GeoDa on Github [Internet]. [cited 2020 Sep 14]. Available from: http://geodacenter.github.io/index.html
10. hiyer09. hiyer09/geopsa_paper [Internet]. 2020 [cited 2020 Sep 14]. Available from: https://github.com/hiyer09/geopsa_paper
11. Friedl MA, Sulla-Menashe D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set] [Internet]. NASA EOSDIS Land Processes DAAC; 2019 [cited 2020 Jul 1]. Available from: https://doi.org/10.5067/MODIS/MCD12Q1.006
12. Center for International Earth Science Information Network - CIESIN - Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia. Global Roads Open Access Data Set, Version 1 (gROADSv1) [Internet]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC); 2013 [cited 2020 Jan 7]. Available from: https://doi.org/10.7927/H4VD6WCT
13. WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University. Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). 5258 https://dx.doi.org/ . /SOTON/WP00645.
14. WorldPop :: Population [Internet]. [cited 2020 Mar 9]. Available from: https://www.worldpop.org/project/categories?id=3
15. Danielson J, Gesch D. GMTED2010 [Internet]. U.S. Geological Survey (USGS) Earth Resources Observation & Science (EROS) Center; 2011. Available from: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation?qt-science_center_objects=0#qt-science_center_objects
16. ESRI. ArcGIS Open Data [Internet]. [cited 2020 Jul 1]. Available from: https://www.esri.com/en-us/arcgis/products/arcgis-open-data

17. Maina J, Ouma PO, Macharia PM, Alegana VA, Mitto B, Fall IS, et al. A spatial database of health facilities managed by the public health sector in sub Saharan Africa. Scientific Data. 2019 Jul 25;6(1):1–8.