Table S1. Steps and statistical tests per process for each type of scenario.

		Applicability per scenario (A,B,C)
Function A	**Data management and adjustment pre-processes**	
Process 1	**"Cleaning" of a database/double records, etc.**	A,B,C
Steps	**Selection**	
	- Select your dataset (big-data, pooled data etc.). - Describe your dataset in details (source, extent, single or combined dataset, type of cancer eg. low or high prevalence etc.).	
	Parsing	
	- Detect any syntax errors. A parser decides whether a string of data is acceptable within the allowed data specification.	
	Data transformation	
	- Map the data from its given format into the format expected by the appropriate application. This includes value conversions or translation functions, as well as normalizing numeric values to conform to minimum and maximum values.	
	Pre-processing	
	- Clean your dataset from multiple entries, missing values and noise.	
	Transformation	
	- Convert (if necessary) the data values from the data format of the cancer registry database to the destination data system (code generation, geo-reference, etc).	
	Data mining	
	- Detect any extremity of the data (outlier/change/deviation detection).	
	Statistical methods for pre-analysis	
	- Analyze the data using the values of mean, standard deviation, range, or clustering algorithms, it is possible for an expert to find values that are unexpected and thus erroneous. Although the correction of such data is difficult since the true value is not known, it can be resolved by setting the values to an average or other statistical value. Statistical methods can also be used to handle missing values in the previous step, which can be replaced by one or more plausible values, which are usually obtained by extensive data augmentation algorithms.	
Process 2	**Linking or integrating data**	A,B,C
Steps	**Bring datasets on a similar format and file type**	
	Use software's commands (e.g. join in GIS) - There are two basic types of joins in GIS tables and geo-data: spatial and relational (based on table IDs). - Spatial join is the key concept in GIS and that is what sets all GIS technologies apart from other relational databases.	
Process 3	**Locating, geo-referencing, editing and recognizing all types of data** (health, social, environmental, satellite images/aerial photos and other quantitative or qualitative data).	A,B,C
Steps	**Import existing digital data in the GIS software** - Direct import of digital data is in most cases the easiest form of digital spatial data conversion. - All software systems provide links to other formats, but the number and functionality of import routines varies between packages. - High-end systems are more likely to provide import functions for a large number of exchange formats. It is also more likely that other data producers will be able to provide GIS data in the native format of the GIS package.	
	Create new entities (e.g. locate cancer data per place of residence) -Use the editing tool and locate cancer cases (x,y coordinates).	
	Add any other type of materials (e.g. photos, images from satellites).	

	Geo-reference and recognize all types of data	
Process 4	**Estimating epidemiological rates, ratios, indexes** etc (e.g. Prevalence, incidence, standardized mortality ratios, etc.) and adjusting data (according to variables that affect the outcomes; e.g. age, sex, socioeconomic).	A,B,C
Steps	**Adjusting data** (according to variables that affect the outcomes; e.g. age, sex, socioeconomic).	
	Calculate incidence or mortality rates -Use standardization techniques to manage any population differences and adjust to age or sex or socioeconomic status (direct or indirect standardization).	
Function B	**Mapping data**	
Process 1	**Visualization modeling** (spatial and/or temporal dimensions to epidemiologic and other data).	
Steps	**Choose whether the representation of cancer data will be in a viewable medium or format**. In GIS, visualization is used to organize spatial data and related information into layers that can be analyzed or displayed as maps, three-dimensional scenes, summary charts, tables, time-based views, and schematics. -Decide if you will visualize your data spatially or temporally too.	A,B,C
	Aggregate cancer data if needed (e.g. for privacy reasons or analysis purposes). -Use distance-weighted averaging such as kernel densities -Use the adaptive buffering	 A,B,C B
Process 2	**Creating maps, videos, interactive maps and animation with real x,y,z dimensions**	
Steps	**Map cancer data using points**	B,C
	Map cancer data using polygons (per municipality or state or other regional level).	A,B,C
	Map cancer data using lines (e.g to illustrate cancer patient flows).	B,C
Process 3	**Exporting descriptive statistics in the form of graphs or maps** (spatial mean and median, standard distance/ellipse, central mean, etc.).	
Steps	**Identify the spatial mean and median, standard distance/ellipse, central mean**	A,B,C
	Measure spatial tendency -The coordinate-wise mean of a point set is the centroid, which solves the same variation problem in the plane (or higher-dimensional Euclidean space) that the familiar average solves on the real line — that is, the centroid has the smallest possible average squared distance to all points in the dataset.	A,B,C
	Measure spatial dispersion to capture the degree to which points in a point set are separated from each other. -Several simple measures of spatial dispersion for a point set can be defined using the covariance matrix of the coordinates of the points. The trace, the determinant, and the largest eigenvalue of the covariance matrix can be used. -A measure of spatial dispersion that is not based on the covariance matrix is the average distance between nearest neighbors.	A,B,C
	Measure spatial autocorrelation to assess and analyze the degree of dependency among observations in a geographic space. -Use the Moran's Index -Use the Geary's C -Use the Getis's G -Use the local Moran's Index	 B,C A,B,C A,B,C A,B,C
	Measure spatial homogeneity -A simple probability model for spatially homogeneous points is the Poisson process in the plane with constant intensity function and the Ripley's K and L functions.	B,C
	Measure spatial stratified heterogeneity to explore whether the within-	B,C

	strata variance is less than the between strata-variance. This reflects the essence of nature, implies potential distinct mechanisms by strata, suggests possible determinants of the observed process, allows the representativeness of observations of the earth, and enforces the applicability of statistical inferences. -Measure it by geographical detector q-statistic	
	Perform spatial interpolation techniques to estimate the variables at unobserved locations in geographic space based on the values at observed locations. -Use the inverse distance weighting (his attenuates the variable with decreasing proximity from the observed location). -Use the Kriging (this interpolates across space according to a spatial lag relationship that has both systematic and random components. It can accommodate a wide range of spatial relationships for the hidden values between observed locations. Kriging provides optimal estimates given the hypothesized lag relationship, and error estimates can be mapped to determine if spatial patterns exist).	A,B,C
Function C	**Spatial/spatio-temporal analysis**	
Process 1	**Identifying relationships, clusters and hot spots**	
Steps	**Use global clustering tests** to evaluate and identify cancer clusters without pinpointing the specific locations of clusters. - Use the Diggle and Chetwynd's bivariate K-function, Mantel-Bailar's test, Potthoff-Whittinghill method (PW). -Use the Moran's I statistic, Cuzick and Edwards's nearest neighbors - Tango's maximized excess events tests (MEET).	 A,B,C B,C A,B,C
	Use local clustering tests for specific small-scale clusters and focused cancer clustering. - Use the Anselin's local (hot spots analysis), using the spatial association (LISA) - Use the Besag-Newell test -Use the Kulldorff's spatial test -Use Diggle's test or Stone's conditional test	 A,B B A,B,C B,C
	Use space-time clustering tests to evaluate and identify cancer clusters in spatio-temporal datasets. -Use the global space-time Knox technique -Use the Diggle's global space-time K-function -Use the extended Kuldorff's spatial test - Use the dynamic continuous-area space-time (DYCAST) system - Use the generalized Bayesian maximum entropy (GBME)	 B A,B,C A,B B,C B
Process 2	**Mapping patterns and trends**	
Steps	**Create choropleth maps** to thematically group regions within an area in proportion to the measurement of the statistical variable being displayed on the map (eg. cancer mortality).	A,B
	Simulate and model disease (Complex adaptive systems theory as applied to spatial analysis suggests that simple interactions among proximal entities can lead to intricate, persistent and functional spatial entities at aggregate levels). -Cellular automata modeling imposes a fixed spatial framework -Agent-based modeling uses entities (agents) that have purposeful behavior and react, interact and modify their environment while seeking outcomes.	 A,B B,C
	Explore spatial interactions -Use gravity models to estimate the flow of people or the roots of disease or the flows or new cancer cases. -Use computational methods such as artificial neural networks to identify spatial interaction relationships among locations by limiting noise in cancer data and	 B,C A,B,C

	increasing the data quality.	
	Perform trend analysis (3-dimensional perspective of data using second-order polynomials).	A,B,C
Process 3	**Predicting future patterns**	
Steps	**Use cokriging** to produce prediction, probability, quantile, standard error, and standard error of indicators maps under the same conditions as for the other kriging methods.	B,C
	Use time-series models	B,C
	Use Bayesian projecting techniques	A,B,C
Process 4	**Identifying new (optimum) locations**	
Steps	**Use multi-criteria analysis** (use model builder in GIS)	A,B,C
	Use buffering techniques and other geo-analyst tools	A,B,C
Process 5	**Identifying risk areas and factors**	
Steps	**Use the multivariable spatial regression** (Either frequentist spatial regression or Bayesian spatial regression).	B,C
	Use spatial autoregressive models - Use a simultaneous autoregressive (SAR) model - Use the geographically weighted regression (GWR)	 B,C A,B,C
	Use the Bayesian regression models -Use the Besag York and Mollié (BYM) model (robust enough and includes fixed covariates along with random effects). -Use the multivariate CAR models for joint spatial modeling of bivariate and multivariate diseases with common risk factors. -Use a combination of models: Bayesian hierarchical modeling in conjunction with Markov Chain Monte Carlo (MCMC) are effective in modeling complex relationships using Poisson-Gamma-CAR, Poisson-lognormal-SAR, or Over dispersed logit models. -Use stochastic /Gaussian processes.	 A,B,C B,C A,B,C B,C