# **Spatial Variation in Ischemic Heart Disease Across Europe**

# **Key Points**

**Question**: What is the spatiotemporal association between national Ischemic Heart Disease (IHD) mortality per 100k citizens and environmental, socioeconomic, and healthcare factors?

**Findings**: In a spatial-lag analysis of 53 countries from 36 continuous years ranging from 1985 to 2020, time starting from 1985 (-), GDP per Capita (-), Population over 65 per 100k (+), Fruits and Vegetable Availability (+), Physicians per 100k (-), and the spatial autoregressive term λ were significant at an α < 0.01 level. The λ of 0.25 outweighed all other predictors by magnitude, suggesting that a 1-standard deviation (SD) average increase in the 4 nearest countries’ IHD rate is associated with a 0.25-SD increase in the target country’s rate.

**Meaning**: Targeted IHD improvements in our study region should focus more on geographic neighborhood prevalence than well-known IHD predictors like GDP per Capita.

# **Abstract**

**Importance**: Cardiovascular diseases (CVDs) remain the leading cause of death globally, with IHD being the most common (Vogel et al., 2021).

**Objective:** To quantify spatiotemporal associations of significant environmental, socioeconomic, and healthcare related predictors of IHD mortality rates.

**Main Outcomes and Measures:** Log-transformed and z-score standardized IHD mortality rate (deaths / 100 k), spatial autoregressive parameter λ.

**Results:** λ=0.259 (SE 0.03; P<0.001), indicating significant spatial clustering of IHD rates. GDPperCapita (β=–0.142; SE 0.032; P<0.001) and time trend (β=–0.027; SE 0.002; P<0.001). Model R²=0.803. Residual diagnostics supported spatial and regression model assumptions.

**Conclusions and Relevance:** After controlling for unobserved country traits and spatial dependence, healthcare and overall development decline from West to East Europe strongly predict higher IHD mortality. Spatial modeling with units of varying size can be adopted more widely in global epidemiology.

# **Introduction**

Currently, it is well known that explanatory variables like GDP per capita and smoking prevalence negatively correlate with heart disease deaths by country and that this relationship is well-conserved globally (Baptista and Quieroz, 2022; Yuyun et al., 2020). Further, according to Tobler’s First Law of Geography, neighboring countries are more likely to exhibit both similar heart disease mortality rates and environmental, socioeconomic, and healthcare variables that correlate with higher rates of heart disease (Miller, 2004). A global analysis to predict IHD using Global Burden of Disease (GBD) data from 2013-2017 for 187 countries revealed that spatial lag and error modeling to account for IHD spatial clustering improved Ordinary Least Squares (OLS) model accuracy (Baptista and Quieroz, 2022). Baptista and Quieroz point out that the spatial clustering trend for IHD and their strongest predictor in all regression models, GDP per capita, empirically appears most prevalent in Europe, our study region. That being said, the authors address that using entire countries as units for spatial modeling has limitations due to the Modifiable Areal Unit Problem (MAUP), which can skew heavier spatial weighting towards large countries like Russia that shares far more borders than the average European country (Wong, 2009; Baptista and Quieroz, 2022). Baptista and Queiroz recognize that subnational heterogeneity among IHD and its predictor variables exists, but they did not subdivide larger countries for the spatial modeling. Instead, they used available country-level indicators and matched these to neighboring countries using spatial weights calculated by Queen Contiguity to honor political boundaries. The authors used OLS, spatial lag, and spatial error models all with an R² between 0.3-0.4 with 4 predictors: GDP per capita, Percent Urbanization, Schooling, Cigarettes per 100k.

The original study objective was to develop an accurate and generalizable model using generalized linear modeling, random forests, and gradient boosting with the predictors variables consistently present across the countries and time period of interest within the World Health Organization (WHO) dataset. Exploratory data analysis revealed that IHD decreased over time and exhibited strong spatial clustering within the study region preserved across the 36 years of interest. Thus, we hypothesized that a spatiotemporal model with a single continuous time variable would improve model diagnostics. Limitations of current spatial analyses of global IHD include sparse data availability, addressing MAUP, and dilution of spatial trends across large geographic areas (Baptista and Quieroz, 2022). As follows, we aim to reproduce a similar country-level spatial analysis of IHD to Baptista and Quieroz specific to our study-region, taking additional measures to address known significant predictors - Percent Urbanization, schooling, cigarettesper100k - being over 75% missing in our larger time-range of 1985-2020. A highly spatially autocorrelated study region, kNN-centroid computed spatial weights matrix, and row-standardization of queen contiguity aim to address current IHD spatial modeling limitations.

A map of europe with red and blue countries/regions

AI-generated content may be incorrect.

A map of europe with different colored countries/regions

AI-generated content may be incorrect.

Figure 1. Study Area projected with ERTS 1989 LAEA. Spatial clustering of average IHD in Europe is preserved over 4 decades: 1980s, 1990s, 2000s, 2010s. The 5 quintiles in each plot quadrant show the decrease in IHD mortality rate since 1980.

# **Methods**

## **Study Design and Data Sources**

This study utilizes two principal datasets: 1) the World Health Organization (WHO) covering demographic, environmental, socioeconomic, and health outcome data across 64 countries from 1949 to 2024 2) the Global Burden of Disease (GBD) study by the Institute for Health Metrics and Evaluation (IHME), which aims to provide global health metrics, details CVD mortality rates from 1990 to 2022. The WHO dataset is notable for its breadth of time and variables but contains considerable missing data, especially before 1985. Conversely, the GBD dataset offers detailed IHD but only from 1990 onwards. We then performed preprocessing to clean, merge, and align the IHD data across different formats and sources. We chose to narrow our analysis to 53 countries of interest from 1985-2020 from the merged dataset for a complete and balanced panel of (year, IHD mortality).

## **Variables**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a spreadsheet

AI-generated content may be incorrect.

ALB AND AUT AZE PRT BLR BEL BIH BGR ESP FRA HRV CYP CZE DNK EST FIN GEO DEU GRC HUN ISL IRL ISR ITA KAZ KGZ LVA LTU LUX MLT MDA MCO MNE NLD MKD NOR POL ROU RUS SMR SRB SVK SVN SWE CHE TJK TUR TKM UKR GBR UZB

Table 1. a,b,c. a. and b. displays the important variables selected by permutation tests before and after missing value imputation, respectively. Both tables are shown prior to log-transformation, z-score standardization, and further variable removal. Table c shows the 53 the countries of interest listed by ISO 3-digit code.

## **Statistical Analysis**

We addressed missing values for important WHO predictors with a hybrid approach: for strongly interdependent health and dietary variables we performed regionally guided iterative imputation to preserve their mutual correlations, while for variables with weaker interdependencies but clear geographic patterns like socioeconomic or system metrics, we used a region‐stratified K-nearest neighbors imputer (Protopapas, 2024). Simpler demographic and uniformly missing variables were filled using robust median imputation to maintain overall distributional trends without overfitting (Protopapas, 2024). After performing log-transformations for right-skewed variables, z-score standardization for all variables, Variance Inflation Factor (VIF) analysis and subsequent removal of variables with high multicollinearity, we further reduced dimensionality by dropping predictors with insignificant Mean Squared Error (MSE) increase following OLS permutation (Protopapas, 2024; Glickman, 2025). Moran’s I was performed for the remaining log-transformed dependent variable and all predictor variables. Given prevalence of MAUP for spatial units of varying size and presence of islands in our dataset, we first built a spatial weights matrix in ArcGIS pro using Queen Contiguity for the 53 countries to obtain an average number of border neighbors (k = 4) to build a balanced spatial weights matrix using kNN for country centroids in R (Chen, 2025; Esri).

Empirical Bayes smoothing flagged countries with total population < 50 000 via countrypops from the gt package in R: AND, MCO, SMR (Iannone et al., 2025). However, smoothing had a negligible effect as expected given insignificant correlation between country population and standardized IHD rate. Both spatial lag and spatial error models were tested like in Baptista and Quieroz 2022 though OLS model residuals yielded less significant spatial autocorrelation (p=0.04), suggesting usage of a spatial lag term that could account for the spill-over of the dependent variable from nearby countries (Anselin, 1988). Finally, conditional-likelihood specification yielded better model accuracy than pooling and random effects due to its ability to control for time-invariant heterogeneity for each country’s intercept (Millo and Piras, 2012).

## Reporting Standards

Software: R 4.4.x, splm, spdep, sf, ggplot2, purrr, dplyr, readr, plm

model\_lag <- spml( formula = as.formula(formula\_str\_time), listw = listw, data = panel\_data, model = "within", lag = TRUE, # include spatial lag of the dependent variable spatial.error = "none", # exclude spatial error term

Within (condition on αᵢ) effects spatial-lag model via spml() was performed using the splm package in R (Millo and Piras, 2012). Likelihood-ratio drop-one tests were performed for variable selection in the spatial lag model (Table 2) (Xenakis, 2024).

# **Results (≈ 900 words)**

## **Diagnostics**

2a,b

A screenshot of a computer program

AI-generated content may be incorrect.

2c,d

A screen shot of a computer

AI-generated content may be incorrect.

2e,f

A screen shot of a computer

AI-generated content may be incorrect.

2g

A table of numbers with black text

AI-generated content may be incorrect.

2h

A screenshot of a computer

AI-generated content may be incorrect.

2i

Moran I test under randomisation

data: resid\_ols

weights: listw

Moran I statistic standard deviate = 1.7345,

p-value = 0.04141

alternative hypothesis: greater

sample estimates:

Moran I statistic Expectation Variance

0.127258119 -0.019230769 0.007132666

2j

Moran I test under randomization

data: resid\_splm

weights: listw

Moran I statistic standard deviate = -1.4042,

p-value = 0.9199

alternative hypothesis: greater

sample estimates:

Moran I statistic Expectation Variance

-0.137107515 -0.019230769 0.007047136

Table 2a-i. 2a-f show the process of the spatial lag model first built with important predictors from OLS. LR drop-one selection subsequently removed UnemploymentRate and HospitalBedPer100k with α < 0.05 though the model with HospitalBedPer100k performed better in R² and AIC shown in table g. alongside pooled spatial lag, random effects spatial lag, and ‘within’ spatial error models. Table h Shows the OLS models for log-transformed predictors with significant skewness > 1 and necessary variable removal for VIF > 5. The second part of table h under "Original Permutation Importance" represents all models ran for the original 8 predictors chosen by permutation importance prior to considering skewness and VIF, some of which scored highly in multiple model diagnostics but violated assumptions of homoscedasticity and normality for linear models and generalized poorly for ensemble models. Table i shows OLS residuals have significant spatial autocorrelation without accounting for spatial lag. Table j shows that Moran’s I on country‐mean residuals are insignificant after adding a spatial lag term.

A map of europe with a plot of results

AI-generated content may be incorrect.

Figure 2a,b,c. a. Map of residuals averaged over all 36 years by country, residuals vs. fitted values. b. Residual vs. fitted plot meets the assumption of homoscedasticity with the exception of outliers. c. QQ plot meets the assumption of normality apart from tails.

## **Descriptive Statistics and Initial Findings**

**Final Spatial-Lag Model Estimates**

y=λ(IT​⊗W)y + (IT ⊗X)β+ (IT ⊗ IN)α + ε,

where:

* y is the (NT)×1 stacked vector of IHD‐mortality over N countries and T years,
* W is the N×N spatial‐weights matrix created with kNN,
* (IT​⊗W) replicates W in a block matrix for each year,
* X is the N×5 matrix of selected covariates: time, GDP, FV, 65+, Phys
  + β is the 5×1 vector of their slopes
* α collects the N unit fixed effects,
* λ is the spatial‐autoregressive coefficient (≈0.259), and
* ε is the NT×1 idiosyncratic error.

Tiling the same spatial weights matrix W along the diagonal for each year pools information across time so that only one λ is estimated (LeSage and Pace, 2009). The equation conditions out α before fitting the model, removing fixed effects before estimation. By subtracting each country’s time-invariant IHD mean α, all non-spatial heterogeneity is removed before β estimation (Bhatt et al., 2017). λ(IT​⊗W)y shrinks residuals in nearby space by pulling each yit towards the 4-neighbor average and replicates the same spatial shrinkage for each year T, smoothing annual fluctuations. As follows, the spatial lag model covariance component focuses on the remaining, spatiotemporally smoothed residual structure, which can lead to simpler covariance forms and better generalization accuracy at the cost of N-1 degrees of freedom shown in Table 2g (Bhatt et al., 2017).

The final spatial lag model reveals a strong positive spatial spillover (λ = 0.259, p < 0.001), meaning that a one-standard-deviation increase in the log-transformed, standardized IHD mortality rate among four neighboring units is associated with a 0.259-SD increase locally. Over time starting from 1985, there is a significant decrease in the log-transformed, standardized IHD mortality rate (β\_time = –0.027, p < 0.001), indicating that IHD mortality rate declines by 0.027 SD each year. One SD increase in log-transformed, standardized GDP per capita is linked to a 0.142 SD decrease in the log-transformed, standardized IHD mortality rate (β\_GDP = –0.142, p < 0.001), supporting the similar global analysis of Baptista and Quieroz. In contrast, a one SD increase in log-transformed, standardized fruit and vegetable availability is associated with a 0.080 SD increase in the log-transformed, standardized IHD mortality rate (β\_FV = 0.080, p < 0.001), suggesting an unexpected positive relationship. A larger share of the population aged 65+ corresponds to increased risk (β\_65+ = 0.110, p < 0.01), reflecting elevated IHD mortality in aging countries. Finally, more physicians per 100 000 inhabitants significantly reduces the log-transformed, standardized IHD mortality rate (β\_phys = –0.084, p < 0.01), highlighting the protective effect of greater medical coverage.

# **Discussion**

The most significant findings of this study were the success of spatial models in improving prediction of IHD across Europe. Current research has focused primarily on time series analysis to predict future levels of disease but focused less on the spatial correlation around the globe (Khan et al., 2020). On the other hand, spatial analysis studies have observed similar patterns in IHD in smaller localities like subnational units within Ethiopia, for example (Alemu et al., 2024). Interestingly, the authors produced a successful Geographic Weighted Regression (GWR) to predict IHD from environmental variables in this Ethiopia-based study with a lower value of IHD spatial autocorrelation calculated by Moran’s I (Alemu et al., 2024). In our research scope of Europe and Western Asia, is it more feasible to predict IHD deaths across larger areas with a higher sample size of nested regions within each country? Or, do larger sample sizes of data for each country mask subnational variation and skew higher levels of spatial clustering between countries than that of Alemu et al?

Nonetheless, one challenge in addition to MAUP when modeling over larger regions in our study instead of subnational units included handling missing values due to the lack of data availability across countries. Many countries disclose varying amounts of information, which made it difficult to determine correlation structures across continuous variables and space with available data from WHO and GBD over the time period of interest from 1985-2020. Some strengths of this study included sophisticated imputation methods and reliable data sources. However, some limitations include relying heavily on missing data imputation methods and a small sample size of countries restricted to our non-comprehsensive study region: some countries, Bosnia and Herzegovina for instance, were dropped due to inconsistent IHD mortality rate availability despite being within the study region boundaries.

Wang et al performed comprehensive spatial modeling of IHD across subnational spatial units in China and suggests testing multiple spatial weights matrices, and our kNN model proved more accurate than the Queen Contiguity model in R², adj R², and AIC, supporting our hypothesis that kNN addressed MAUP better than Queen Contiguity used in global IHD Spatial Models with units of varying size and border contiguity (Baptista and Queiroz, 2022). Further, this study supports the findings of Wang et al, that different spatial weight matrices can lead to drastically different conclusions, given that (λ≈0.25; p < 0.001) was far higher and more significant in the kNN spatial lag model while the Queen Contiguity model (λ≈0.09; p < 0.01) assigned higher coefficient values to GDP and other continuous predictors, suggesting that more IHD spatial autocorrelation was preserved in kNN.

The observed significant positive association between fruit and vegetable availability and ischemic heart disease mortality contrasts with the prevailing literature, which generally reports an inverse relationship. (Oude Griep et al., 2023). This suggests that the WHO FruitsVeggiesAvailability variable masks the effect of other variable(s) in our study. For example, our preliminary analysis of the WHO data revealed a strong negative correlation between FruitsVeggiesAvailability / LiteracyRate (-0.39) and FruitsVeggiesAvailability / HospitalBedper100k (-0.42), which would better describe the positive relationship between FruitsVeggiesAvailability and IHD explained by the model given Baptista and Quieroz’s GBD ‘Schooling’ variable and our negatively correlated healthcare viability predictors HospitalBedper100k and Physiciansper100k.

The strengths of this study include an emphasis on the spatial lag parameter λ = 0.259 and the single continuous time parameter. For IHD policy intervention, the study results suggest that a country’s four nearest neighbors’ IHD rates in Europe provide a better indicator of IHD mortality than any continuous variable, even widely recognized GDPpercapita. Moreover, the significant time term β\_time shows the consistent decline in IHD that is recognized globally from 1985-2020 and conserved in our study region. However, recent studies note that the long-term global decline in CVD mortality is decreasing due to increases in known predictors like obesity, suggesting that future CVD models for inference should take this concerning inflection point into account (Murray, 2023).

The practical implications of this alignment with current research is that more attention should be given to spatial models as a tool to predict disease in target areas over large-scale data. It is already evident that many explanatory variables can accurately predict disease outcomes for IHD (Baptista and Queiroz, 2022). Our study reaffirms the consensus that public health interventions should focus on IHD mortality rate trends of a country’s neighbors and less developed countries (Baptista and Quieroz, 2022). Further, this study echoes the limitations of important variations within countries remaining unaddressed from Baptista and Quieroz. Thus, future studies should investigate subnational units as global data becomes more available in the information age, enabling multi-scale spatial analyses (Cui et al. 2022).

# **Conclusion**

We advocate for the integration of spatial modeling in global health and data science research as we found a strong positive spatial spillover (λ = 0.259, p < 0.001), indicating that neighboring countries’ IHD rates exert a larger influence on a country’s mortality than traditional continuous predictors. Our study was interdisciplinary by default, and exploratory data analysis proved that an ensemble of spatial, socioeconomic, environmental, and healthcare variables all significantly correlate with higher levels of heart disease, suggesting that continuous collaboration and data sharing among disciplines will lead to the most accurate predictions of heart disease deaths.

# **References**

Alemu, Yohannes M., et al. "Nationwide Geospatial Analysis to Identify Variations in Primary Cardiovascular Risk in Ethiopia." *Journal of Primary Care & Community Health*, vol. 15, 2024, doi:10.1177/21501319241288312.

Anselin, Luc. “Spatial Dependence in Regression Error Terms.” *Spatial Econometrics: Methods and Models*, edited by Luc Anselin, Springer Netherlands, 1988, pp. 100–18. *Springer Link*, <https://doi.org/10.1007/978-94-015-7799-1_8>

Baptista, Emerson Augusto, and Bernardo Lanza Queiroz. “Spatial Analysis of Cardiovascular Mortality and Associated Factors around the World.” *BMC Public Health*, vol. 22, no. 1, Aug. 2022, p. 1556. *BioMed Central*, <https://doi.org/10.1186/s12889-022-13955-7>.

Bhatt, Samir, et al. “Improved Prediction Accuracy for Disease Risk Mapping Using Gaussian Process Stacked Generalization.” *Journal of The Royal Society Interface*, vol. 14, no. 134, Sept. 2017, p. 20170520. *DOI.org (Crossref)*, <https://doi.org/10.1098/rsif.2017.0520>.

Cui, Wencong, et al. "Analyzing Spatial Variations of Heart Disease and Type-2 Diabetes: A Multi-Scale Geographically Weighted Regression Approach." *Computational Urban Science*, vol. 2, 2022, article number 34, doi:10.1007/s43762-022-00059-6.

Di Maria, Alissia and Christescu, Ioanna, “APCOMP209a Final Project Milestone 3: Exploratory Data Analysis.” *APCOMP209a*, Harvard University, 2024.

Di Maria, Alissia and Christescu, Ioanna, “APCOMP209a Final Project Milestone 4: Missing Value Imputation Before Modeling.” *APCOMP209a*, Harvard University, 2024.

ESRI. "World Countries Generalized." *ArcGIS Hub*, ESRI,<https://hub.arcgis.com/datasets/esri::world-countries-generalized/about>. Accessed 9 December 2024.

Khan, Moien AB, et al. "Global Epidemiology of Ischemic Heart Disease: Results from the Global Burden of Disease Study." *Cureus*, vol. 12, no. 7, 2020, e9349, doi:10.7759/cureus.9349.

Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo J, Brevoort K, Roy O (2025). *gt: Easily Create Presentation-Ready Display Tables*. R package version 1.0.0.9000, [https://gt.rstudio.com](https://gt.rstudio.com/)

LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

Miller, Harvey J. "Tobler’s First Law and Spatial Analysis." *Annals of the Association of American Geographers*, vol. 94, no. 2, 2004, pp. 284–289, doi:10.1111/j.1467-8306.2004.09402005.x.

Millo, Giovanni, and Gianfranco Piras. “Splm: Spatial Panel Data Models in R.” *Journal of Statistical Software*, vol. 47, Apr. 2012, pp. 1–38. *www.jstatsoft.org*, <https://doi.org/10.18637/jss.v047.i01>.

Morehouse, Kirsten, “APCOMP209a Final Project Milestone 2: Data Cleaning.” *APCOMP209a*, Harvard University, 2024.

Morehouse, Kirsten, “APCOMP209a Final Project Milestone 5: ARIMA and SEM analysis.” *APCOMP209a*, Harvard University, 2024.

Murray, Christopher J. L. “The Global Burden of Disease Study at 30 Years.” *Nature Medicine*, vol. 28, no. 10, Oct. 2022, pp. 2019–26. *DOI.org (Crossref)*, <https://doi.org/10.1038/s41591-022-01990-1>.

Oude Griep, Linda M., et al. “Worldwide Associations of Fruit and Vegetable Supply with Blood Pressure from 1975 to 2015: An Ecological Study.” *BMJ Nutrition, Prevention & Health*, vol. 6, no. 1, June 2023, pp. 28–38. *DOI.org (Crossref)*, <https://doi.org/10.1136/bmjnph-2022-000455>.

Protopapas, Pavlos. "Random Forests." *APCOMP209a*, Harvard University, 2024. Lecture.

Taliotis, Alyssa Mia, “APCOMP209a Final Project Milestone 1: Project Proposal.” *APCOMP209a*, Harvard University, 2024.

Taliotis, Alyssa Mia, “APCOMP209a Final Project Milestone 4: Model Optimization and Selection.” *APCOMP209a*, Harvard University, 2024.

Vogel, Birgit, et al. "The Lancet Women and Cardiovascular Disease Commission: Reducing the Global Burden by 2030." *The Lancet*, vol. 397, no. 10292, 2021, pp. 2385-2438.

Wang, W., et al. “[Influences of using different spatial weight matrices in analyzing spatial autocorrelation of cardiovascular diseases mortality in China].” *Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi*, vol. 42, no. 8, Aug. 2021, pp. 1437–44. *PubMed*, <https://doi.org/10.3760/cma.j.cn112338-20201102-01293>.

Wong, D. W. “Modifiable Areal Unit Problem.” *International Encyclopedia of Human Geography*, edited by Rob Kitchin and Nigel Thrift, Elsevier, 2009, pp. 169–74. *ScienceDirect*, <https://doi.org/10.1016/B978-008044910-4.00475-2>.

Xenakis, James. “Mixed Models.” *STAT139,* Harvard University, 2024. Lecture.

Yuyun, Matthew Fomonyuy, et al. "Cardiovascular Diseases in Sub-Saharan Africa Compared to High-Income Countries: An Epidemiological Perspective." *Global Heart*, vol. 15, no. 1, 2020.

References from Conversations with Dr. Chen.

This study is an extension of a large-scale analysis of the WHO dataset performed for APCOMP209a by Kent Codding, Alissia Di Maria, Ioanna Christescu, Kirsten Morehouse, and Alyssa Mia Taliotis. Each contributor played a large role in this modeling project, and this paper focuses on the geospatial analysis performed in ArcGIS Pro by Kent Codding. The other four contributors did extensive work regarding data acquisition, data cleaning, nonspatial exploratory data analysis, and model optimization and selection, excluding geographic weighted regression. Their work is referenced throughout the paper.

You can list the top three journals in your mind and their writing requirements here

**JAMA**

**Meta-analysis**

These manuscripts are systematic, critical assessments of literature and data sources pertaining to clinical topics, emphasizing factors such as cause, diagnosis, prognosis, therapy, or prevention, and that includes a statistical technique for quantitatively combining the results of multiple studies that measure the same outcome into a single pooled or summary estimate. All articles or data sources should be searched for and selected systematically for inclusion and critically evaluated, and the search and selection process should be described in the manuscript. The specific type of study or analysis, population, intervention, exposure, and tests or outcomes should be described for each article or data source. The data sources should be as current as possible, ideally with the search having been conducted within several months of manuscript submission. Authors of reports of meta-analyses of clinical trials should submit the [PRISMA flow diagram and checklist](https://www.equator-network.org/reporting-guidelines/prisma). Authors of meta-analyses of observational studies should submit the [MOOSE checklist](https://jamanetwork.com/journals/jama/fullarticle/192614). Follow [EQUATOR Reporting Guidelines](https://www.equator-network.org/).

A structured abstract is required; for more information, see instructions for preparing [Abstracts for Meta-analysis](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecAbstractsforMeta-analyses). A list of 3 Key Points is required (see guidance on preparing [Key Points](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecKeyPoints)). Maximum length: 3000 words of text (not including abstract, tables, figures, acknowledgments, references, and online-only material), with no more than a total of 5 tables and/or figures and no more than 50-75 references. The subtitle should include the phrase "A Meta-analysis." To read more about meta-analyses, see the [*AMA Manual of Style*](https://academic.oup.com/amamanualofstyle/book/27941/chapter/207572072#382796687).

**Original Investigation**

These reports typically include randomized trials (see [Clinical Trial](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecStudyTypeGroup1)), intervention studies, cohort studies, case-control studies, epidemiologic assessments, other observational studies, surveys with high response rates (see [Reports of Survey Research](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecReportsofSurveyResearch)), cost-effectiveness analyses and decision analyses (see [Reports of Cost-effectiveness Analyses and Decision Analyses](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecReportsofCost-effectivenessAnalysesandDecisionAnalyses)), and studies of screening and diagnostic tests (see also [Reports of Diagnostic Tests](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecReportsofDiagnosticTests)). Each manuscript should clearly state an objective or hypothesis; the design and methods (including the study setting and dates, patients or participants with inclusion and exclusion criteria and/or participation or response rates, or data sources, and how these were selected for the study); the essential features of any interventions; the main outcome measures; the main results of the study; a discussion section placing the results in context with the published literature and addressing study limitations; and the conclusions and relevant implications for clinical practice or health policy. Data included in research reports must be original and should be as timely and current as possible (see [Timeliness of Data](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecTimelinessofData)). Follow [EQUATOR Reporting Guidelines](https://www.equator-network.org/).

A structured abstract is required; for more information, see instructions for preparing [Abstracts for Reports of Original Data](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecAbstractsforReportsofOriginalData). A list of 3 Key Points is required (see guidance on preparing [Key Points](https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecKeyPoints)). Maximum length: 3000 words of text (not including abstract, tables, figures, acknowledgments, references, and online-only material) with no more than a total of 5 tables and/or figure