

Methods for Estimating the Global Burden of Cerebrovascular Diseases

Gregory A. Roth^{a, b} Catherine O. Johnson^b Grant Nguyen^b Mohsen Naghavi^b
Valery L. Feigin^c Christopher J.L. Murray^b Mohammed H. Forouzanfar^b
Theo Vos^b

^aDivision of Cardiology, Department of Medicine, and ^bInstitute for Health Metrics and Evaluation, University of Washington, Seattle, Wash., USA; ^cNational Institute for Stroke and Applied Neurosciences, Faculty of Health and Environmental Sciences, Auckland University of Technology, Auckland, New Zealand

Key Words

Ischemic stroke · Hemorrhagic stroke · Population health

Abstract

The Global Burden of Disease (GBD) study is a long-standing effort to report consistent and comprehensive measures of disease burden for the world. In this paper, we describe the methods used to estimate the global burden of stroke for the GBD 2013 study. Pathologic subtypes of stroke are modeled separately for two mutually exclusive and exhaustive categories: (1) ischemic stroke and (2) hemorrhagic and other non-ischemic strokes. Acute and chronic strokes are estimated separately. The GBD 2013 study has incorporated large amounts of new data on stroke death rates, incidence and case fatality. Disease modeling methods have been updated to better integrate mortality and incidence data. Future efforts will focus on incorporating data on the regional variation in severity of disability. Stroke remains a new area for disease modeling. A better understanding of stroke incidence, mortality and severity, and how it varies among countries, can help guide priority setting and improve health policy related to this important condition. © 2015 S. Karger AG, Basel

Introduction

The Global Burden of Disease (GBD) study is a long-standing effort to report consistent and comprehensive measures of disease burden for the world [1]. In 2010, the GBD study produced the first global estimates of death, disability and overall health loss due to ischemic and hemorrhagic strokes [2]. The GBD 2013 study represents the most recent update to this work, supported by a collaboration of over 1,000 investigators in more than 50 countries to estimate global incidence, prevalence and mortality of over 300 diseases. In addition, the study estimates summary measures for health including years of life lost compared with the best observed lifespan, years lived with disability (YLDs) and disability-adjusted life years. Health loss is then attributed to more than 70 risk factors in the domains of behavioral, metabolic and environmental/occupational exposures. Estimation for the GBD study makes use of custom software to handle enormous amounts of input data and advanced methods in computational epidemiology. Results are produced separately by sex and 5-year age categories, for the years 1990–2013. In this paper, we describe the methods used to estimate the global burden of stroke for the GBD 2013 study.

Methods

Case Definitions and Disease Categories

The GBD study has adopted the World Health Organization definition for stroke as ‘rapidly developing clinical signs of focal disturbance of cerebral function lasting more than 24 h or leading to death with no apparent cause other than that of vascular origin’ [3, 4]. Transient ischemic attacks, epidural and subdural hemorrhage, trauma, malignancy and infection as causes of neurological deficits are excluded. Pathologic subtypes of stroke are modeled separately for two mutually exclusive categories: (1) ischemic stroke and (2) hemorrhagic and other non-ischemic strokes, including sub-arachnoid hemorrhage. Acute stroke and chronic stroke are estimated separately for each subtype. Acute stroke is defined as lasting from symptom onset to 30 days while chronic stroke lasts from 31 days until death.

Mortality

The approach to estimating cause-specific mortality for the GBD study had been previously described [5, 6]. For the GBD 2013 study, a significant amount of new mortality data was added. For all causes of death, the number of site-years of mortality data increased from 2,798 to 5,039 vital registration sources and from 486 to 538 verbal autopsy sources. Substantial amounts of new data were included from the countries of China, Russia and Turkey. Table 1 shows the number of site-years of stroke mortality data used for the GBD 2013 study. Table 2 shows the International Classification of Diseases version 10 (ICD-10) codes used to define ischemic and hemorrhagic or other stroke. As in the GBD 2010 study, non-underlying or ill-defined causes of death were redistributed to one of these two pathological subtypes using statistical methods [7, 8]. For the GBD 2013 study, deaths coded as due to transient ischemic attack (ICD-10 codes G45 and G46) were coded as ischemic stroke and deaths coded as due to non-ruptured aneurysms (ICD code I67.0) were coded as hemorrhagic stroke.

Global and regional mortality due to stroke was estimated in the Cause of Death Ensemble model (CODEm), an ensemble model method developed for the GBD study. CODEm evaluates many different combinations of regression models, two types of cause of death data (rates and cause fractions) and a wide range of country-level data used as covariates [9]. An ensemble is then created based on out-of-sample predictive validity measures and the weighting of each sub-model. Stroke was modeled separately by pathologic subtypes (ischemic, hemorrhagic and other non-ischemic) and also separately for all strokes together, using country-level data as additional covariates for estimates across all included countries. Table 3 shows the covariates selected by the ten CODEm models that performed best on the out-of-sample predictive validity tests. Age-, sex-, country- and year-specific death rates from the stroke sub-type models were adjusted to fit the rates estimated from the all-stroke model and, subsequently, with cause of death rates for all-cause mortality. Years of life lost are estimated as the product of the count of deaths and the GBD standard life expectancy at age of death [10].

Systematic Review of Published Data

A systematic review of the scientific literature reporting local or national population-based data on incidence, prevalence or case fatality due to stroke was performed to update the results of the GBD 2010 study through the year 2013. Two formal queries were used in the PubMed database. Query 1 was the following:

Table 1. Site-years of data used to estimate mortality due to stroke, by pathologic stroke subtype and data source type

Stroke type	Source type	Site-years
Ischemic	Surveillance	481
Ischemic	Vital registration	3,976
Hemorrhagic or other	Surveillance	492
Hemorrhagic or other	Vital registration	3,975

Table 2. ICD-10 codes used for cause of death estimation

Stroke subtype	ICD-10 codes
Ischemic strokes	G45–G46.8, I63–I63.9, I65–I66.9, I67.2, I67.3, I67.5, I67.6, I69.3–I69.398
Hemorrhagic or other non-ischemic strokes	I60–I61.9, I62.0–I62.03, I67.0, I67.1, I67.7, I69.0, I69.198, I69.2–I69.298

Table 3. Country-level data selected by the ten most heavily weighted CODEm models

Tobacco use
Smoking prevalence
Cumulative cigarette consumption
Metabolic
Mean body mass index
Mean systolic blood pressure
Socioeconomic
Years of educations per capita
Income per capita
Health system access summary score
Dietary
Alcohol intake in liters per capita
Nut and seed consumption in kilocalories per capita
Pulse and legume consumption in kilocalories per capita
Omega 3 polyunsaturated fat consumption in kilocalories per capita
Milk consumption in kilocalories per capita
Animal fat consumption in kilocalories per capita
Fruit consumption in kilocalories per capita
Omega 6 polyunsaturated fat consumption in kilocalories per capita
Whole grain consumption in kilocalories per capita
Red meat consumption in kilocalories per capita

((stroke[Mesh]) AND (prevalence[Title/Abstract] OR incidence [Title/Abstract]) AND ('2010'[Date – Publication]: '3000'[Date – Publication]) AND (hasabstract[text] AND Humans[Mesh] AND middle age[MeSH])). Query 2 was the following: ((hasabstract[text] AND Humans[Mesh] AND middle age[MeSH]) OR 21) AND ((hemorrhagic stroke/epidemiology[Mesh] OR hemorrhagic

Table 4. Unpublished stroke registry data included in GBD 2013

Danish Stroke Registry 2009
Israel National Acute Stroke Registry 2004–2010
Joinville, Brazil Stroke Registry 2010–2011
Ludhiana, India Population-Based Stroke Registry, Christian Medical College and Hospital Ludhiana 2010–2013
Mumbai, India Contributory Health Service Scheme Stroke Dataset 2013
Nalchik, Russia Stroke Register 2003–2004
Swedish Stroke Register 2012
Tandil, Argentina Program for the Epidemiological Evaluation of Stroke (PREVISTA) 2013–2014
Ulyanovsk, Russia Stroke Registry 2004–2006
WHO MONICA Cerebrovascular Event Data for China, Denmark, Finland, Hungary, Italy, Lithuania, Poland, Russia, Serbia, and Sweden 1982–1997
Yakutsk, Russia Stroke Registry 2002–2004
Krasnodar, Russia Stroke Registry 1997–2000

stroke/mortality[Mesh]) AND (prevalence[Title/Abstract] OR incidence[Title/Abstract]) AND ('2010'[Date - Publication]: '3000'[Date - Publication]) AND (hasabstract[text] AND Humans[Mesh] AND middle age[MeSH]).

Unpublished data from hospital and community-based stroke registries were identified via collaborating stroke experts in multiple countries (table 4). The complete list of citations identified and those used for modeling are shown in the online supplementary appendix table (for all online suppl. material, see www.karger.com/doi/10.1159/000441083).

Incidence and Prevalence

First-Ever Acute Stroke and Chronic Stroke

Disease modeling was performed using DisMod, a Bayesian meta-regression software package, that applies the approach of epidemiologic state-transition modeling to estimate consistent parameters of incidence, prevalence and mortality. The meta-regression component allows quantification of systematic bias between data sources using different methods and coefficients for the covariates that can help predict values for geographies with sparse or no data. A layered hierarchy of random effects on 7 global super-regions, 21 GBD regions and countries further assists in estimating values for all countries and time periods. DisMod methods have been reported in detail, elsewhere [11, 12]. Figure 1 is a flow chart that shows the steps in estimating stroke prevalence by acuity and pathologic subtype that are now described in more detail.

DisMod was used to separately model epidemiologic parameters for three categories of acute stroke (fig. 1, left side): first ever acute ischemic, first ever acute hemorrhagic or other stroke, and first ever all stroke (including both ischemic and hemorrhagic strokes). These models of acute stroke used data on acute stroke incidence and 30-day case fatality. Review of the published scientific literature provided 1,974 rows of age-sex-country-stroke subtype-specific data for the model while the registry data provided 1,894 rows of data, and population surveys provided 1,555 rows of data. The model for acute ischemic stroke used 2,272 rows of age-sex-country-specific incidence data and 446 rows of excess mortality data. The model for acute hemorrhagic stroke used 1,760

rows of age-sex-country-specific incidence data and 378 rows of excess mortality data. The model for all acute stroke used 6,866 rows of age-sex-country-specific incidence data and 270 rows of excess mortality data. The model for chronic stroke utilized 1,846 rows of age-sex-country-specific data on the prevalence on chronic stroke. Data were obtained by covering 74% of countries for ischemic stroke and 73% of countries for hemorrhagic stroke. Data on the case fatality proportions in the first 30 days were transformed into excess mortality rates using the following equation:

$$\text{excess mortality rate} = \ln \frac{1 - \text{case fatality proportion}}{1/12}.$$

We assumed the duration for acute stroke to be 30 days but no remission from chronic stroke. We applied an assumption within the model that the slope of incidence and excess mortality always increased with age. Ischemic and hemorrhagic or other stroke estimates for incidence were rescaled to the envelope of incidence from a model of first-ever stroke of any type. We then multiplied incidence by one minus the 30-day case fatality proportion estimated from the excess mortality output of the model, using the reverse of the equation described above to estimate the incidence of chronic stroke, effectively producing an estimate of the incidence of surviving to 30 days after a stroke. This incidence of chronic stroke was used in a new DisMod model along with prevalence data from surveys and literature and excess mortality from literature to model epidemiologic parameters for chronic stroke. Results of the estimation process, including citations of all data used for modeling, are available at <http://vizhub.healthdata.org/epi/>.

Stroke mortality data from vital registration and verbal autopsy do not differentiate between deaths that occurred during an acute episode of stroke and deaths from chronic stroke. Therefore, we used the acute stroke death rates output from DisMod (as excess mortality rate divided by prevalence) and subtracted these values from the total stroke deaths rates produced by CODEm to estimate the chronic stroke death rate. This estimated chronic stroke death rate was used to inform the DisMod model for chronic stroke. This was done separately for ischemic stroke, hemorrhagic stroke and all strokes combined.

We modeled chronic stroke as a single model for all stroke types in order to be able to use prevalence data on stroke not specified by subtype. We assume that after the acute phase, mortality rates from ischemic and hemorrhagic stroke were similar. This approach is supported by analyses of linked data between the Perth Stroke study and the Western Australian vital registration system [13]. The results of the combined chronic stroke model were split back into pathologic stroke subtypes using the ratios of ischemic to hemorrhagic incidence rates after surviving the acute episode.

All Acute Stroke

In a second pathway (fig. 1, right side), DisMod was used to model the epidemiologic parameters for all episodes of acute stroke. The same modeling strategy as described above for first-ever acute stroke was employed but, in this case, we relied on data that reflected the incidence of all stroke events, inclusive of recurrent events. This was done to estimate prevalence for all acute stroke episodes, regardless of whether or not they were first-ever strokes, to correctly estimate the burden of disability experienced in the first 30 days after all acute strokes for a given population.

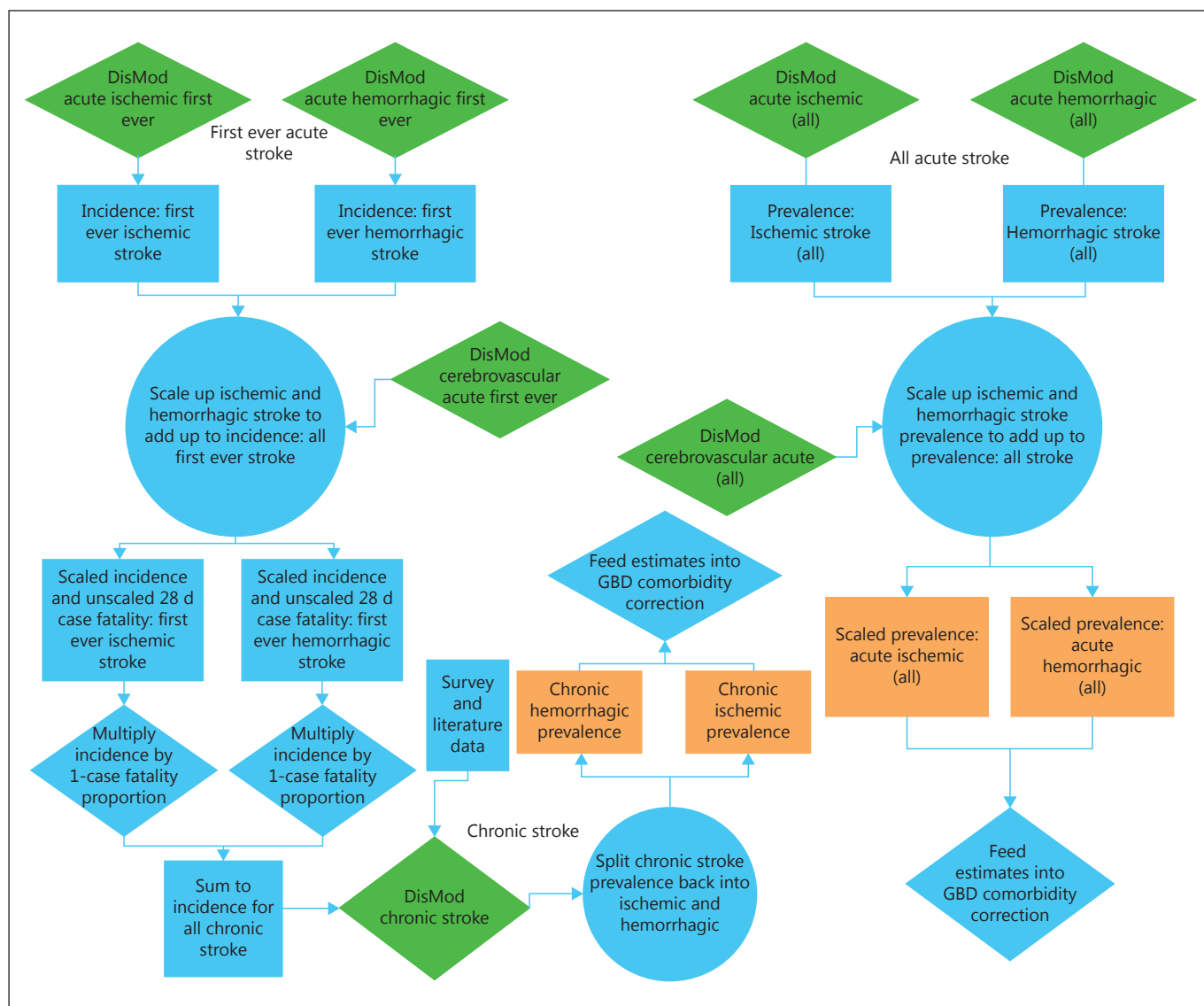


Fig. 1. Relationship between disease models used to estimate stroke prevalence by acuity and pathologic subtype.

Disability

For GBD 2013, disability weights were derived for 230 unique health states using a combined analysis of data from 5 country surveys and an open-access internet survey conducted as part of the GBD 2010 study and 4 more recent European surveys [14, 15]. In these surveys, respondents were presented with lay descriptions for random pairs of health states and asked to make a judgement which of two individuals in either of the two health states is the healthier. The proportion of respondents for any given pair indicating that one of the two health states is healthier determined the relative distance between severity values for either health state. Additional population health equivalence questions were posed for a subset of the 230 health states to help anchor the values from the pair-wise comparisons onto a scale between zero (no health loss) and one (complete health loss or equivalence to death). Five health

states of increasing severity for stroke were included in the surveys (table 5).

The distribution of severity of stroke by these 5 levels was based on an analysis of the Medical Expenditure Panel Survey in the United States. Survey respondents with stroke were identified from ICD-9 codes and self-reported reasons for encounters with the health system. Each respondent was followed for a 2-year period with new respondents added annually. Twice over the 2-year follow-up period, respondents were asked to report questions from the SF-12, a generic measure of health-related quality of life [16]. Using convenience samples of the Institute for Health Metrics and Evaluation staff and participants of GBD training workshops who filled in SF-12 based on lay descriptions for a subset of 60 GBD health states, we derived a crosswalk between SF-12 answers and the GBD disability weight values. Using regression methods with

Table 5. Stroke health states, lay descriptions and disability weight values in GBD 2013

Health state	Lay description	Disability weight
Stroke, long-term consequences, mild	Has some difficulty in moving around and some weakness in one hand, but is able to walk without help	0.019
Stroke, long-term consequences, moderate	Has some difficulty in moving around, and in using the hands for lifting and holding things, dressing and grooming	0.069
Stroke, long-term consequences, moderate plus cognition problems	Has some difficulty in moving around, in using the hands for lifting and holding things, dressing and grooming and in speaking. The person is often forgetful and confused	0.314
Stroke, long-term consequences, severe	Is confined to bed or a wheelchair, has difficulty speaking and depends on others for feeding, toileting and dressing	0.544
Stroke, long-term consequences, severe plus cognition problems	Is confined to bed or a wheelchair, depends on others for feeding, toileting and dressing and has difficulty speaking, thinking clearly and remembering things	0.578

dummies for each condition reported by the Medical Expenditure Panel Survey respondents, we isolated the amount of disability that we could attribute to each condition rather than any comorbid disease. The distribution of the amount of disability in GBD disability weight values that was attributed to stroke was then used to bin all cases of stroke into 6 categories: no disability from stroke and the 5 levels of stroke severity for which we had weights assuming thresholds at zero and the midpoints between the 5 DW values. For acute stroke, we assumed no proportion without disability.

YLDs are calculated as the product of the prevalence of stroke cases in each of the severity levels by the appropriate disability weight and a correction for comorbidity with any of the other disease sequelae estimated in GBD using microsimulation methods that have been previously described [10]. DALYs are the sum of YLLs and YLDs.

Limitations

Several limitations remain for the methods used to estimate stroke in GBD 2013. Data remain limited from many regions of the world. In particular, we relied on severity data specific to the United States and therefore we did not account for regional variation in chronic stroke disability related to access to treatment and rehabilitation services. We assumed that the distribution of stroke severity is the same for acute and chronic stroke. However, our approach to disability does adjust disability to account for comorbidities and account for the proportion that remain asymptomatic following stroke.

While vital registration death data do not differentiate between acute and chronic stroke mortality, our approach to using death data reflects the incidence of acute stroke and is a plausible method for estimating death for these two categories of stroke.

By modeling all stroke types and first-ever stroke separately, and partly with different data sources, we may introduce inconsistencies between the estimates for the prevalence of acute and chronic stroke. However, this approach allows us to use all available data, including those from studies that did not assess stroke status as incident versus recurrent.

Conclusion and Future Directions

Disease modeling for stroke is a new field with little existing research upon which to rely. We have developed a method that makes use of all available population-level information on the burden of stroke to estimate the global health lost to cerebrovascular disease. Future efforts should focus on three areas. More data should be collected from under-represented regions of the world. Severity distributions should reflect regional variation in disability, utilizing real data such as the modified Rankin score. The complexity of our approach should be addressed in order to decrease the need for scaling of intermediate results. Better stroke-specific models will contribute to improved measurement for this important condition that accounts for a large amount of health lost to non-communicable diseases. A better understanding of stroke incidence, mortality and severity, and how it varies between countries, can help guide priority setting and improve health policy related to this important condition.

Funding

The Bill and Melinda Gates Foundation.

Disclosure Statement

The authors have nothing to disclose.

References

- Ezzati M, Lopez AD, Rodgers AA, Murray CJ: Comparative Quantification of Health Risks. Geneva, World Health Organization, 2004.
- Feigin VL, Forouzanfar MH, Krishnamurthi R, Mensah GA, Connor M, Bennett DA, Moran AE, Sacco RL, Anderson L, Truelsen T, O'Donnell M, Venketasubramanian N, Barker-Collo S, Lawes CM, Wang W, Shinohara Y, Witt E, Ezzati M, Naghavi M, Murray C: Global and regional burden of stroke during 1990–2010: findings from the Global Burden of Disease study 2010. *Lancet* 2014;383:245–254.
- Global Burden of Disease Stroke Expert Group, Bennett DA: Methodology of the global and regional burden of stroke study. *Neuroepidemiology* 2012;38:30–40.
- O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, Rangarajan S, Islam S, Pais P, McQueen MJ, Mondo C, Damasceno A, Lopez-Jaramillo P, Hankey GJ, Dans AL, Yusuf K, Truelsen T, Diener HC, Sacco RL, Ryglewicz D, Czlonkowska A, Weimar C, Wang X, Yusuf S: Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010;376:112–123.
- Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne SD, Dorsey ER, Driscoll T, Duber H, Ebel B, Erwin PJ, Espindola P, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D, Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G, Kassebaum N, Keren A, Khoo JP, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, et al: Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease study 2010. *Lancet* 2012;380:2095–2128.
- GBD 2013 Mortality and Causes of Death Collaborators: Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease study 2013. *Lancet* 2015;385:117–171.
- Ahern RM, Lozano R, Naghavi M, Foreman K, Gakidou E, Murray CJ: Improving the public health utility of global cardiovascular mortality data: the rise of ischemic heart disease. *Popul Health Metr* 2011;9:8.
- Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R: Algorithms for enhancing public health utility of national causes-of-death data. *Popul Health Metr* 2010;8:9.
- Foreman KJ, Lozano R, Lopez AD, Murray CJ: Modeling causes of death: an integrated approach using CODEm. *Popul Health Metr* 2012;10:1.
- Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brughla TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, et al: Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease study 2010. *Lancet* 2012;380:2163–2196.
- Katzenellenbogen JM, Vos T, Somerford P, Begg S, Semmens JB, Codde JP: Excess mortality rates for estimating the non-fatal burden of stroke in western Australia: a data linkage study. *Cerebrovasc Dis* 2010;30:57–64.
- Salomon JA, Vos T, Hogan DR, et al: Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease study 2010. *Lancet* 2012;380:2129; 2013;381:628.
- Haagsma JA, de Noordhout CM, Polinder S, Vos T, Havelaar AH, Cassini A, Devleeschauwer B, Kretzschmar ME, Speybroeck N, Salomon JA: Assessing disability weights based on the responses of 30,660 people from four European countries. *Popul Health Metr* 2015;13:10.
- Cheak-Zamora NC, Wyrwich KW, McBride TD: Reliability and validity of the SF-12v2 in the medical expenditure panel survey. *Qual Life Res* 2009;18:727–735.