Figure S1: The age-standardized point prevalence of ischemic stroke in 2021 for the 21 Global Burden of Disease regions, by sex. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

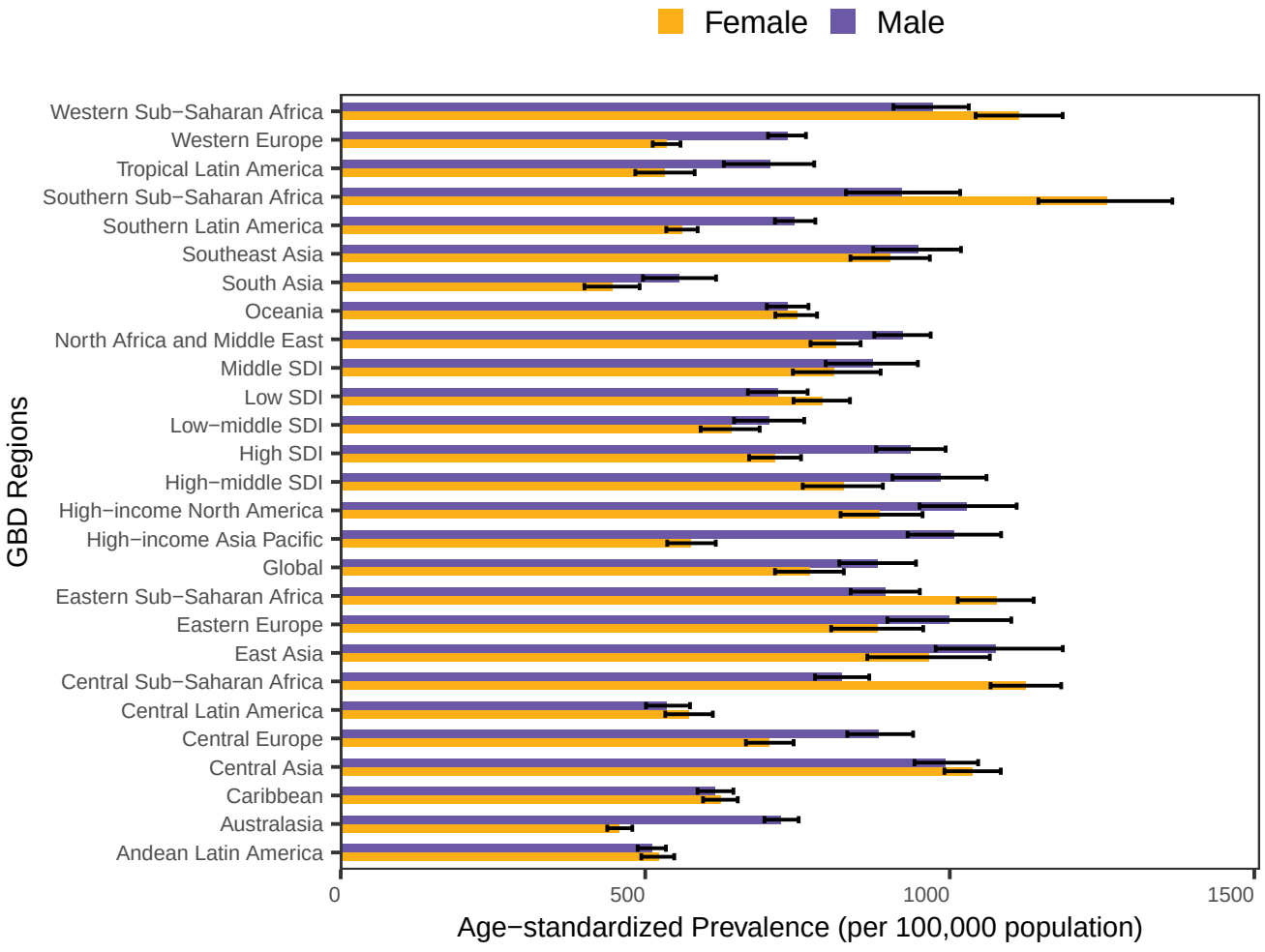

Female
Male
Western Sub−Saharan Africa
Western Europe
Tropical Latin America
Southern Sub−Saharan Africa
Southern Latin America
Southeast Asia
South Asia
Oceania
North Africa and Middle East
Middle SDI
Low SDI
Low−middle SDI
High SDI
High−middle SDI
High−income North America
High−income Asia Pacific
Global
Eastern Sub−Saharan Africa
Eastern Europe
East Asia
Central Sub−Saharan Africa
Central Latin America
Central Europe
Central Asia
Caribbean
Australasia
Andean Latin America
GBD Regions
0
500
1000
1500
Age−standardized Prevalence (per 100,000 population)

Figure S2: The age-standardized incidence of ischemic stroke in 2021 for the 21 Global Burden of Disease regions, by sex. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

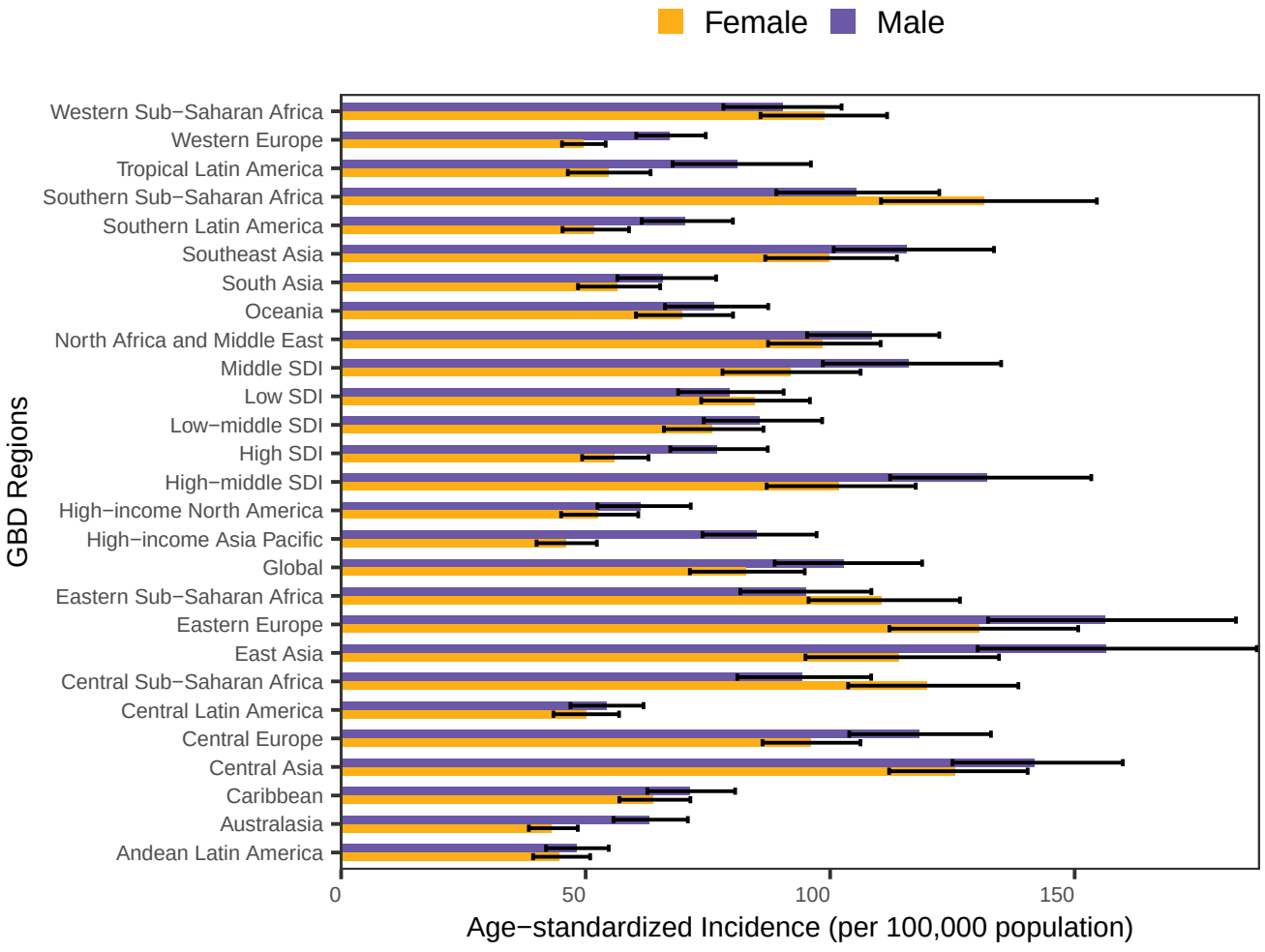

Female
Male
Western Sub−Saharan Africa
Western Europe
Tropical Latin America
Southern Sub−Saharan Africa
Southern Latin America
Southeast Asia
South Asia
Oceania
North Africa and Middle East
Middle SDI
Low SDI
Low−middle SDI
High SDI
High−middle SDI
High−income North America
High−income Asia Pacific
Global
Eastern Sub−Saharan Africa
Eastern Europe
East Asia
Central Sub−Saharan Africa
Central Latin America
Central Europe
Central Asia
Caribbean
Australasia
Andean Latin America
GBD Regions
0
50
100
150
Age−standardized Incidence (per 100,000 population)

Figure S3: The age-standardized death rates of ischemic stroke in 2021 for the 21 Global Burden of Disease regions, by sex. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

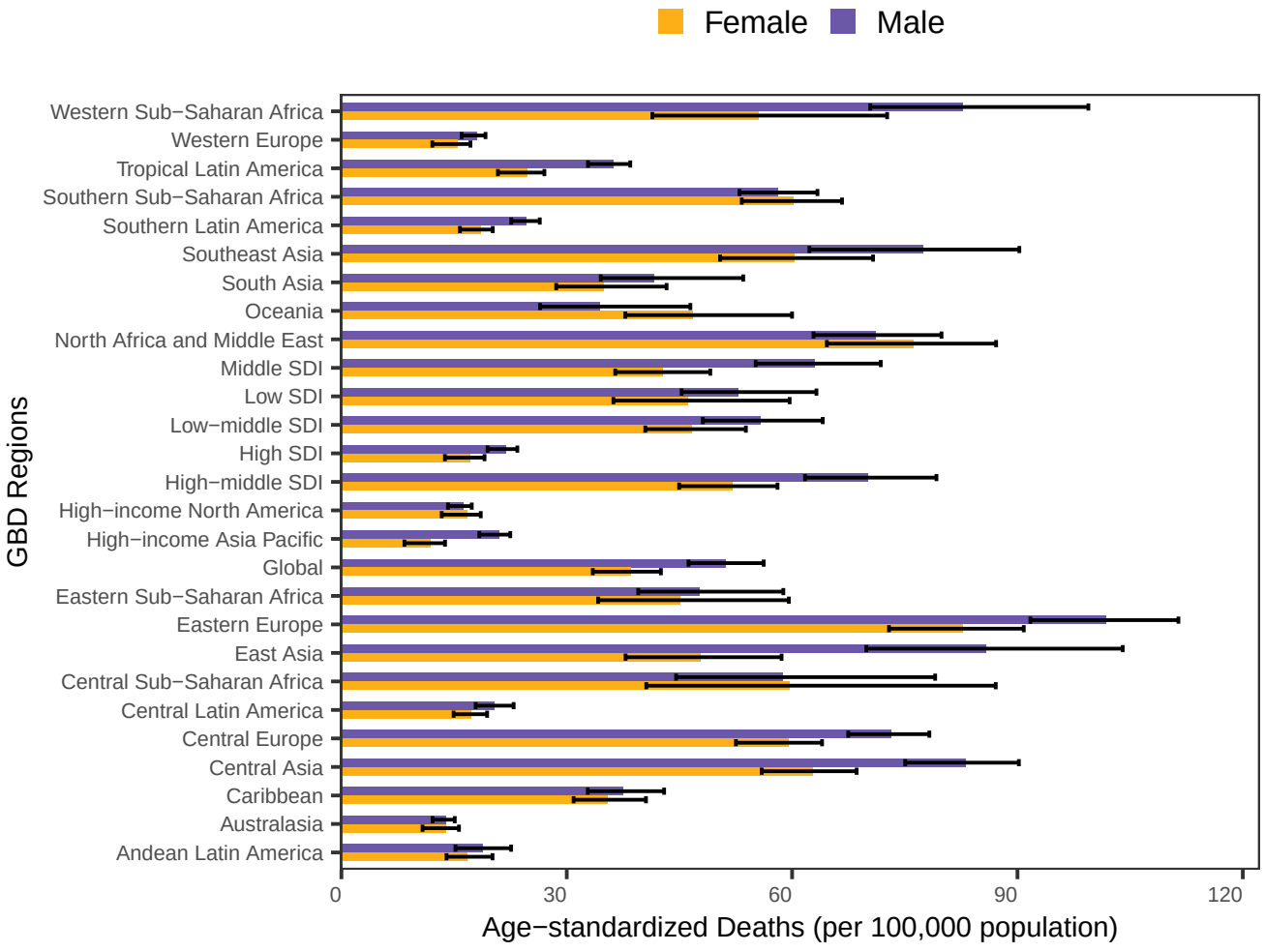

Female
Male
Western Sub−Saharan Africa
Western Europe
Tropical Latin America
Southern Sub−Saharan Africa
Southern Latin America
Southeast Asia
South Asia
Oceania
North Africa and Middle East
Middle SDI
Low SDI
Low−middle SDI
High SDI
High−middle SDI
High−income North America
High−income Asia Pacific
Global
Eastern Sub−Saharan Africa
Eastern Europe
East Asia
Central Sub−Saharan Africa
Central Latin America
Central Europe
Central Asia
Caribbean
Australasia
Andean Latin America
GBD Regions
0
30
60
90
120
Age−standardized Deaths (per 100,000 population)

Figure S4: The age-standardized DALYs of ischemic stroke in 2021 for the 21 Global Burden of Disease regions, by sex.

DALYs = disability-adjusted life years (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

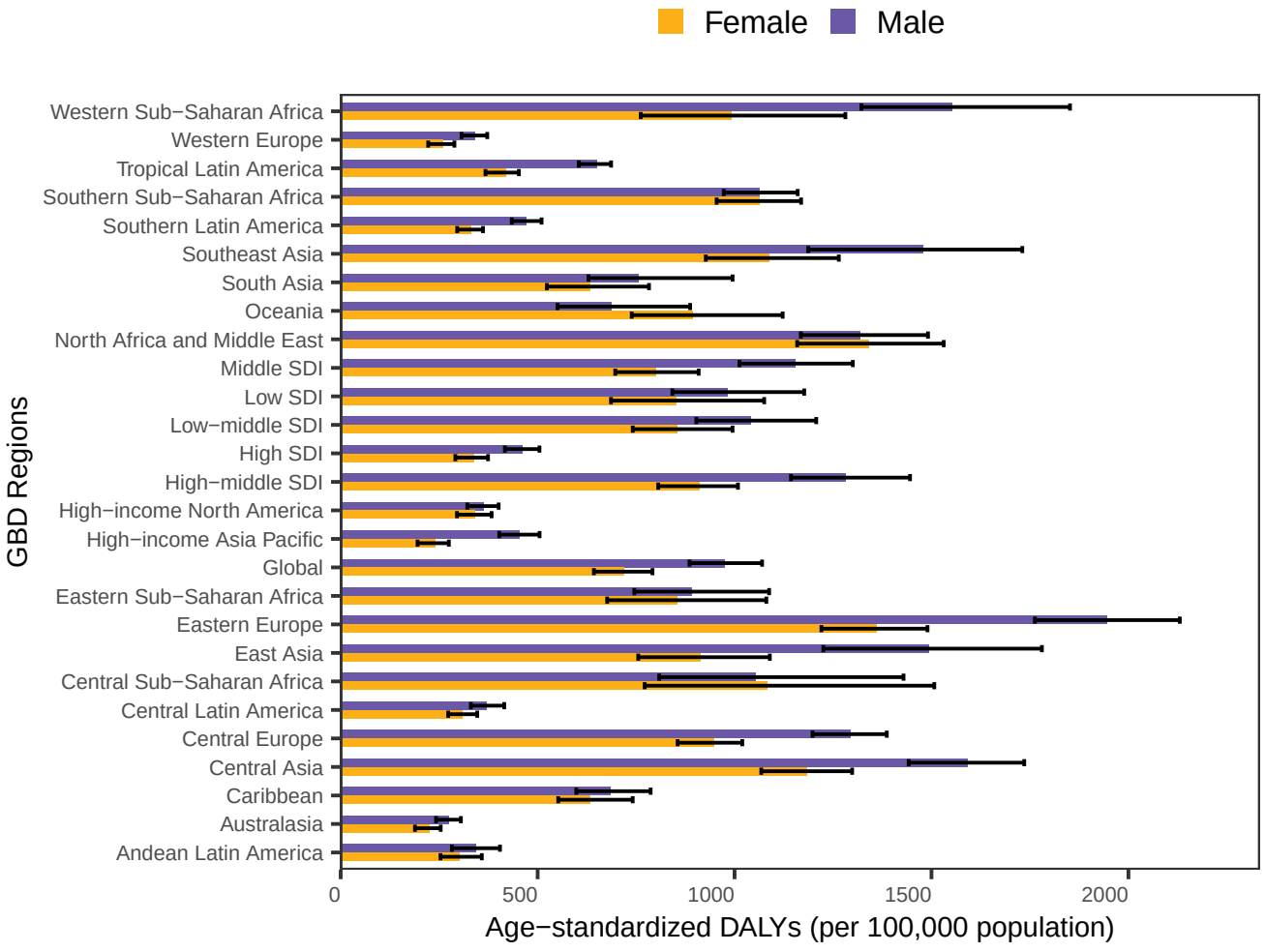

Female
Male
Western Sub−Saharan Africa
Western Europe
Tropical Latin America
Southern Sub−Saharan Africa
Southern Latin America
Southeast Asia
South Asia
Oceania
North Africa and Middle East
Middle SDI
Low SDI
Low−middle SDI
High SDI
High−middle SDI
High−income North America
High−income Asia Pacific
Global
Eastern Sub−Saharan Africa
Eastern Europe
East Asia
Central Sub−Saharan Africa
Central Latin America
Central Europe
Central Asia
Caribbean
Australasia
Andean Latin America
GBD Regions
0
500
1000
1500
2000
Age−standardized DALYs (per 100,000 population)

Figure S5: The percentage change in the age-standardized point prevalence of ischemic stroke from 1990 to 2021 for the 21 Global Burden of Disease regions, by sex. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

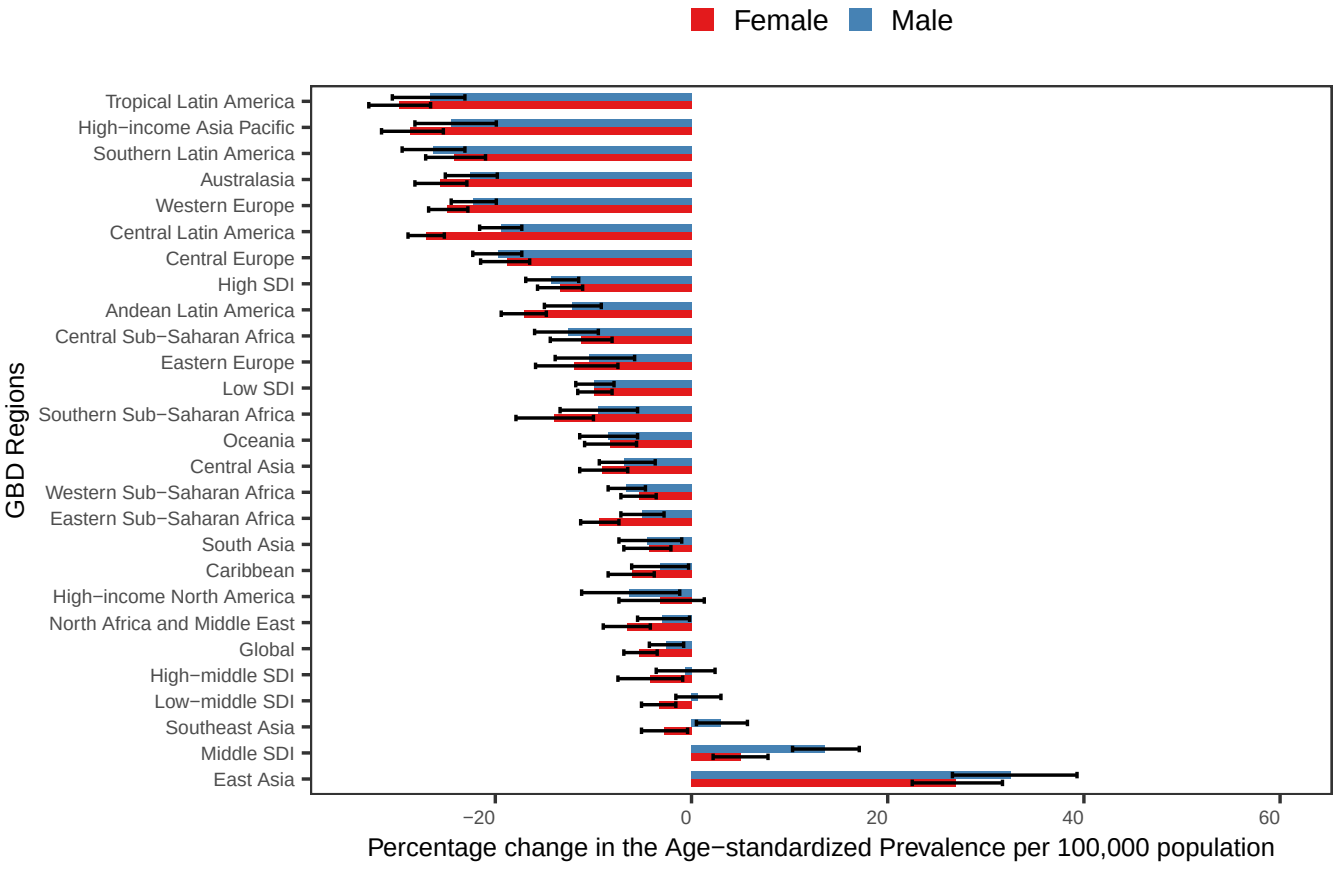

Female
Male
Tropical Latin America
High−income Asia Pacific
Southern Latin America
Australasia
Western Europe
Central Latin America
Central Europe
High SDI
Andean Latin America
Central Sub−Saharan Africa
Eastern Europe
Low SDI
Southern Sub−Saharan Africa
Oceania
Central Asia
Western Sub−Saharan Africa
Eastern Sub−Saharan Africa
South Asia
Caribbean
High−income North America
North Africa and Middle East
Global
High−middle SDI
Low−middle SDI
Southeast Asia
Middle SDI
East Asia
GBD Regions
−20
0
20
40
60
Percentage change in the Age−standardized Prevalence per 100,000 population

Figure S6: The percentage change in the age-standardized incidence of ischemic stroke from 1990 to 2021 for the 21 Global Burden of Disease regions, by sex. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

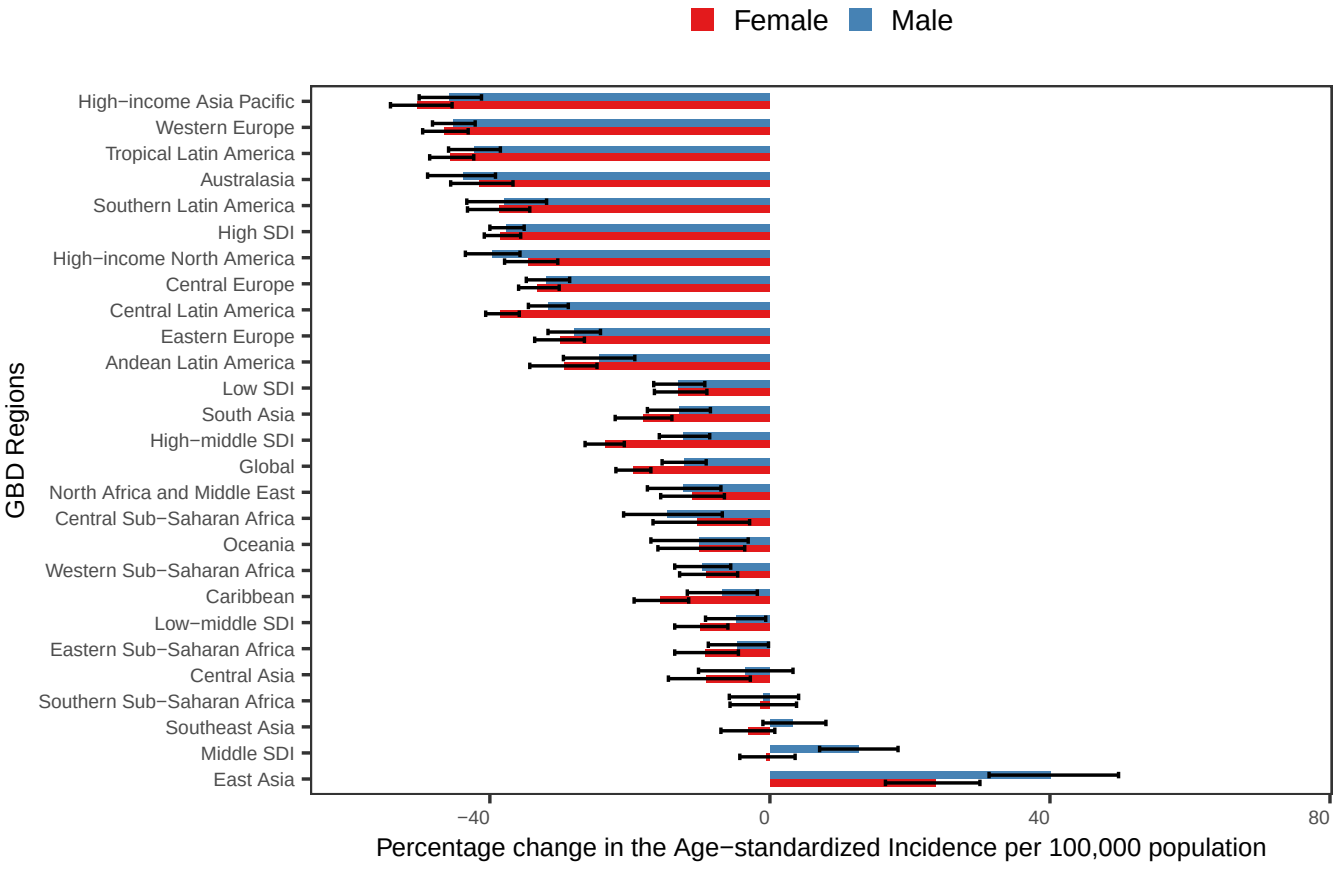

Female
Male
High−income Asia Pacific
Western Europe
Tropical Latin America
Australasia
Southern Latin America
High SDI
High−income North America
Central Europe
Central Latin America
Eastern Europe
Andean Latin America
Low SDI
South Asia
High−middle SDI
Global
North Africa and Middle East
Central Sub−Saharan Africa
Oceania
Western Sub−Saharan Africa
Caribbean
Low−middle SDI
Eastern Sub−Saharan Africa
Central Asia
Southern Sub−Saharan Africa
Southeast Asia
Middle SDI
East Asia
GBD Regions
−40
0
40
80
Percentage change in the Age−standardized Incidence per 100,000 population

Figure S7: The percentage change in the age-standardized death rate of ischemic stroke from 1990 to 2021 for the 21 Global Burden of Disease regions, by sex. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

SDI = Sociodemographic Index, GBD = Global Burden of Disease

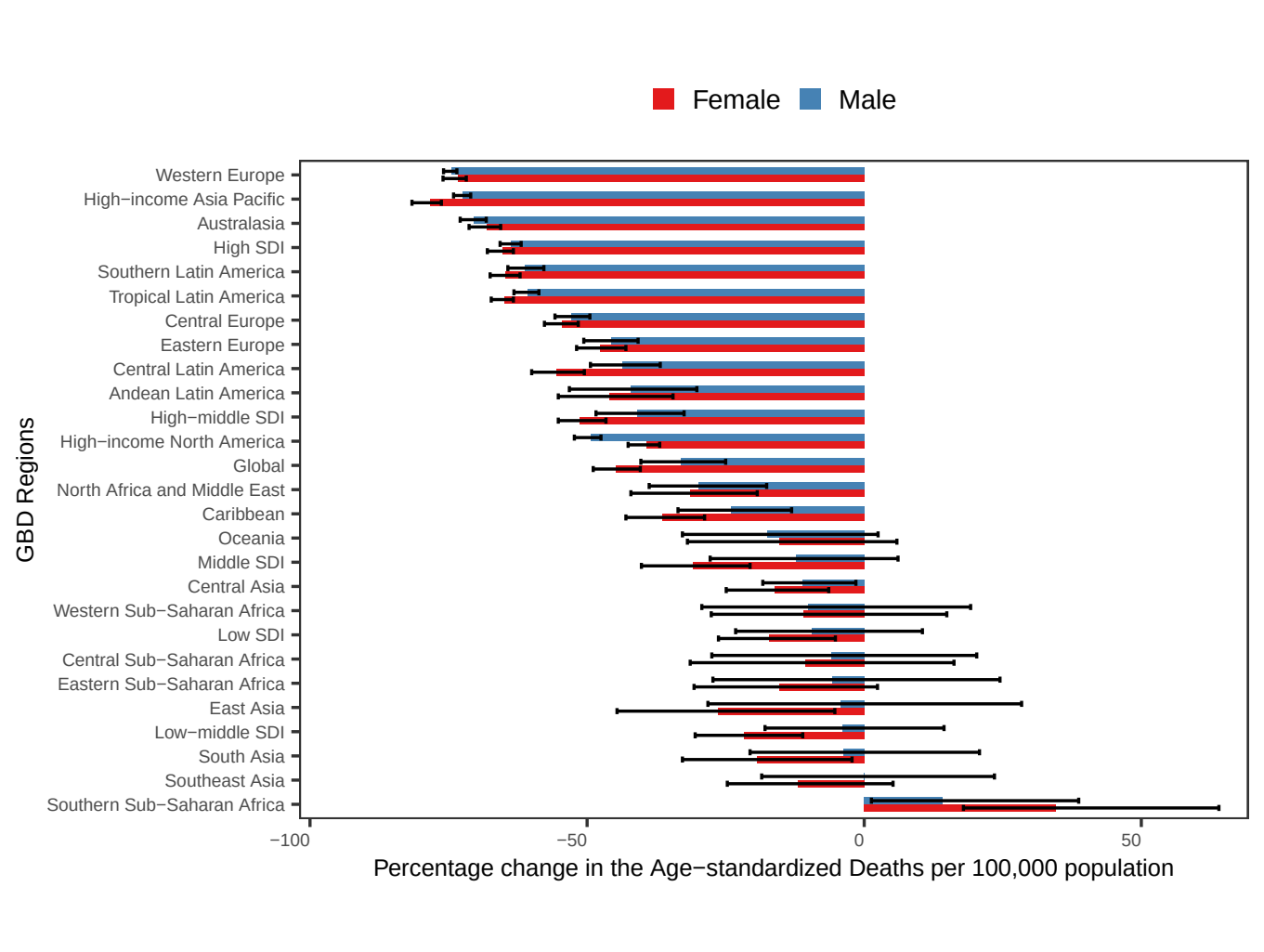

Female
Male
GBD Regions
Western Europe
High−income Asia Pacific
Australasia
High SDI
Southern Latin America
Tropical Latin America
Central Europe
Eastern Europe
Central Latin America
Andean Latin America
High−middle SDI
High−income North America
Global
North Africa and Middle East
Caribbean
Oceania
Middle SDI
Central Asia
Western Sub−Saharan Africa
Low SDI
Central Sub−Saharan Africa
Eastern Sub−Saharan Africa
East Asia
Low−middle SDI
South Asia
Southeast Asia
Southern Sub−Saharan Africa
−100
−50
0
50
Percentage change in the Age−standardized Deaths per 100,000 population

Figure S8: The percentage change in the age-standardized DALYs of ischemic stroke from 1990 to 2021 for the 21 Global Burden of Disease regions, by sex.

DALYs = disability-adjusted life years, SDI = Sociodemographic Index, GBD = Global Burden of Disease (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

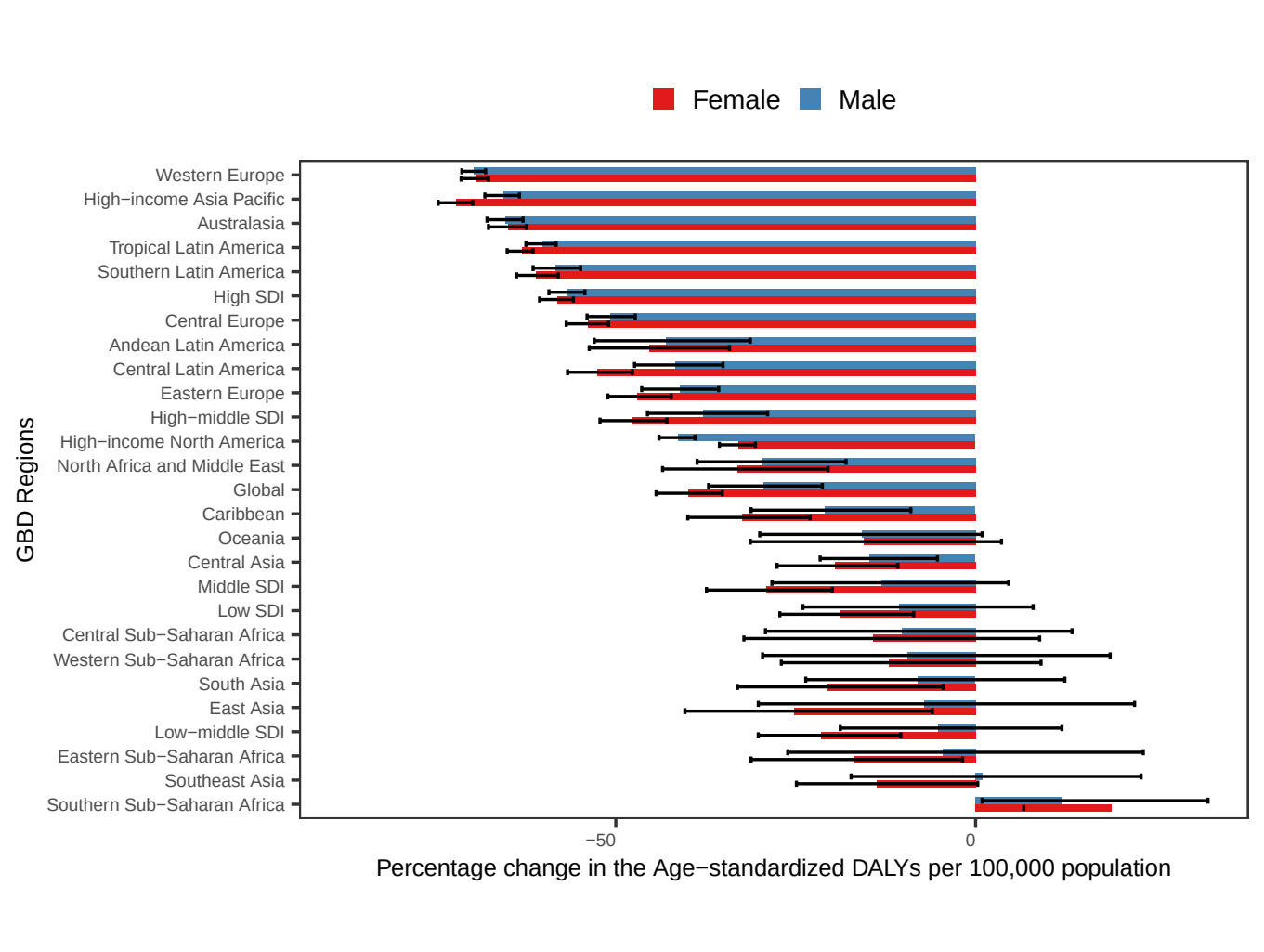

Female
Male
Western Europe
High−income Asia Pacific
Australasia
Tropical Latin America
Southern Latin America
High SDI
Central Europe
Andean Latin America
Central Latin America
Eastern Europe
High−middle SDI
High−income North America
North Africa and Middle East
Global
Caribbean
Oceania
Central Asia
Middle SDI
Low SDI
Central Sub−Saharan Africa
Western Sub−Saharan Africa
South Asia
East Asia
Low−middle SDI
Eastern Sub−Saharan Africa
Southeast Asia
Southern Sub−Saharan Africa
GBD Regions
−50
0
Percentage change in the Age−standardized DALYs per 100,000 population

Figure S9: Age-standardized incidence of ischemic stroke per 100,000 population in 2021, by country. (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

ASIR = Age-Standardized Incidence Rate

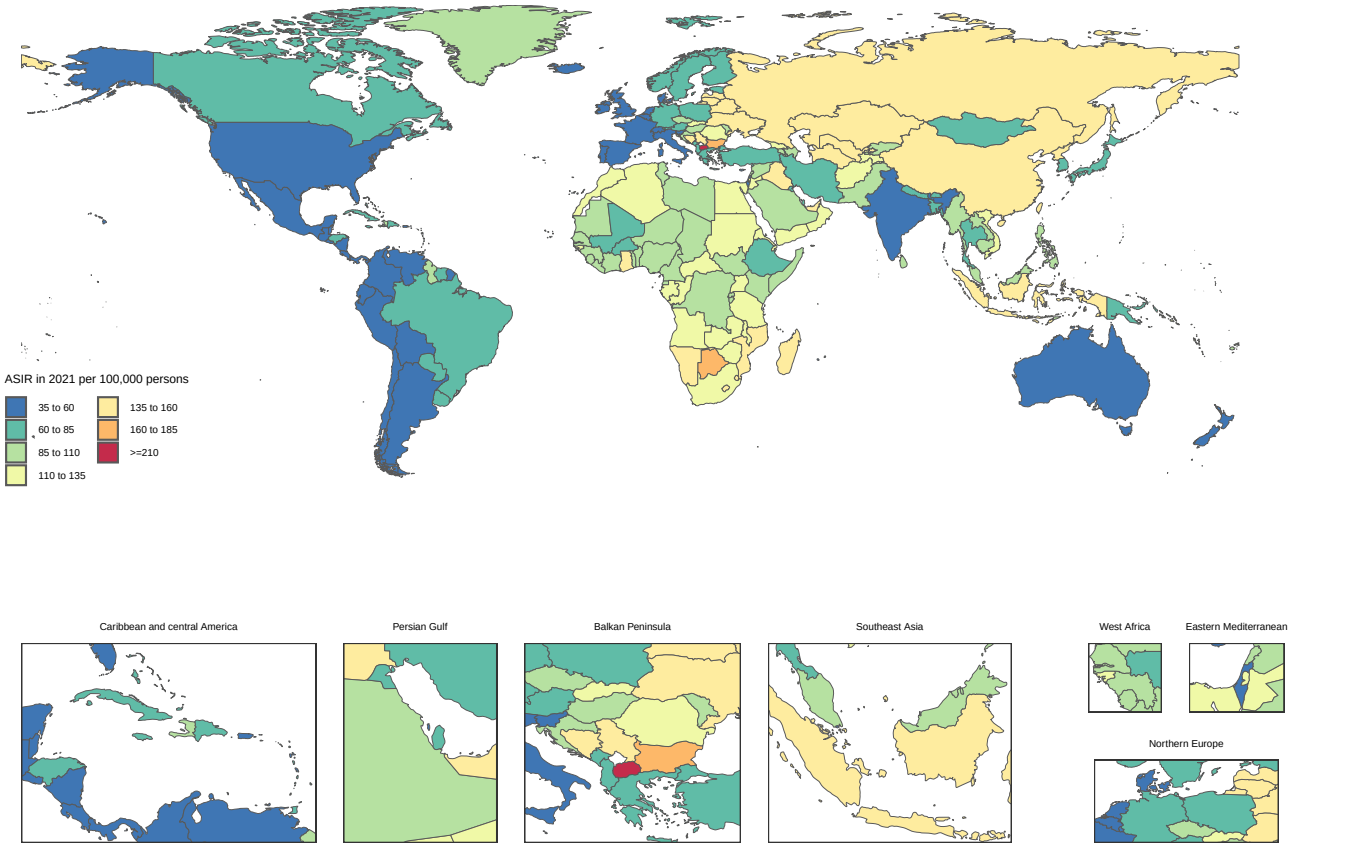

ASIR in 2021 per 100,000 persons
35 to 60
60 to 85
85 to 110
110 to 135
135 to 160
160 to 185
>=210
Caribbean and central America
Persian Gulf
Balkan Peninsula
Southeast Asia
West Africa
Eastern Mediterranean
Northern Europe

Figure S10: Age-standardized death rate of ischemic stroke per 100,000 population in 2021, by country. (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

ASDR = Age-Standardized Death Rate

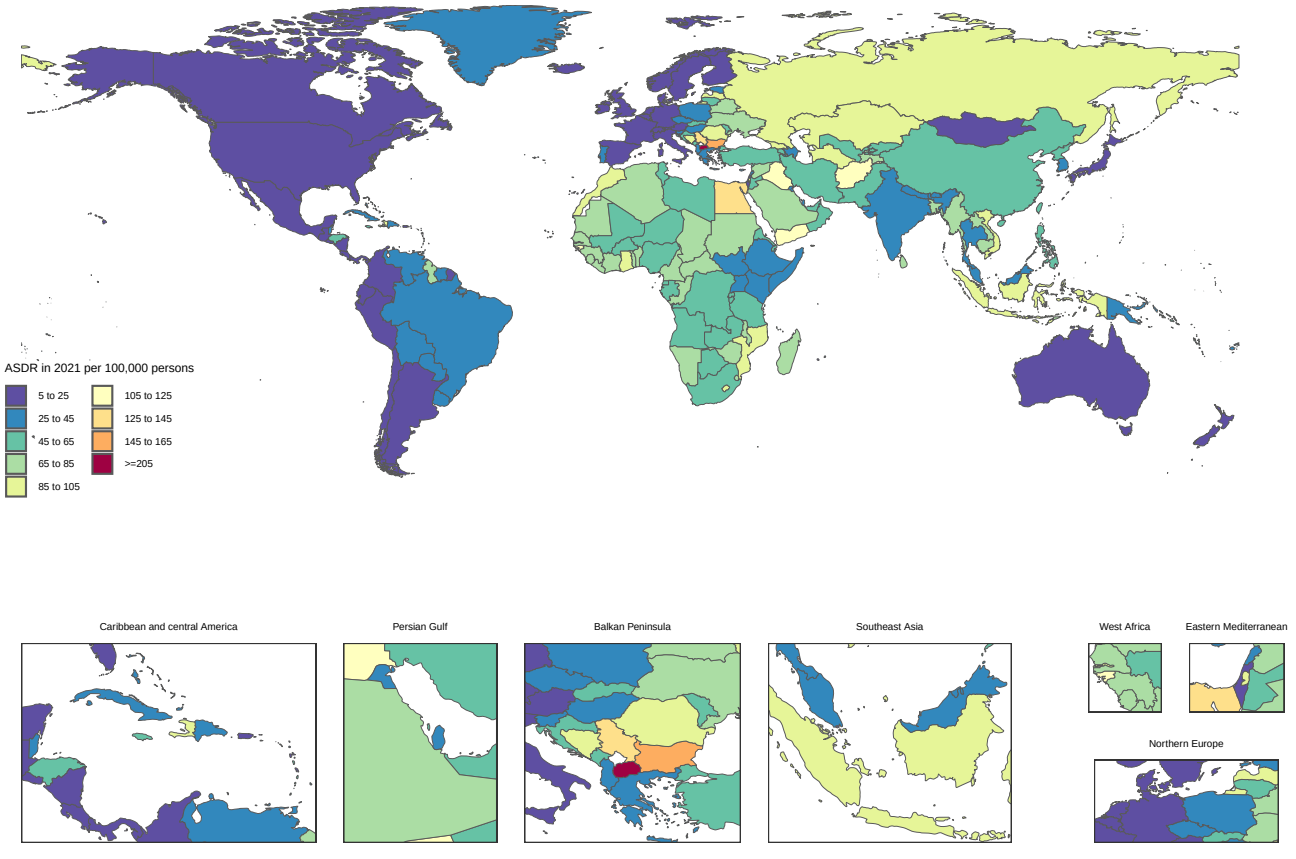

ASDR in 2021 per 100,000 persons
5 to 25
25 to 45
45 to 65
65 to 85
85 to 105
105 to 125
125 to 145
145 to 165
>=205
Caribbean and central America
Persian Gulf
Balkan Peninsula
Southeast Asia
West Africa
Eastern Mediterranean
Northern Europe

Figure S11: Age-standardized DALYs of ischemic stroke per 100,000 population in 2021, by country.

DALYs = disability-adjusted life years (Generated from data available from https://ghdx.healthdata.org/gbd-results-tool).

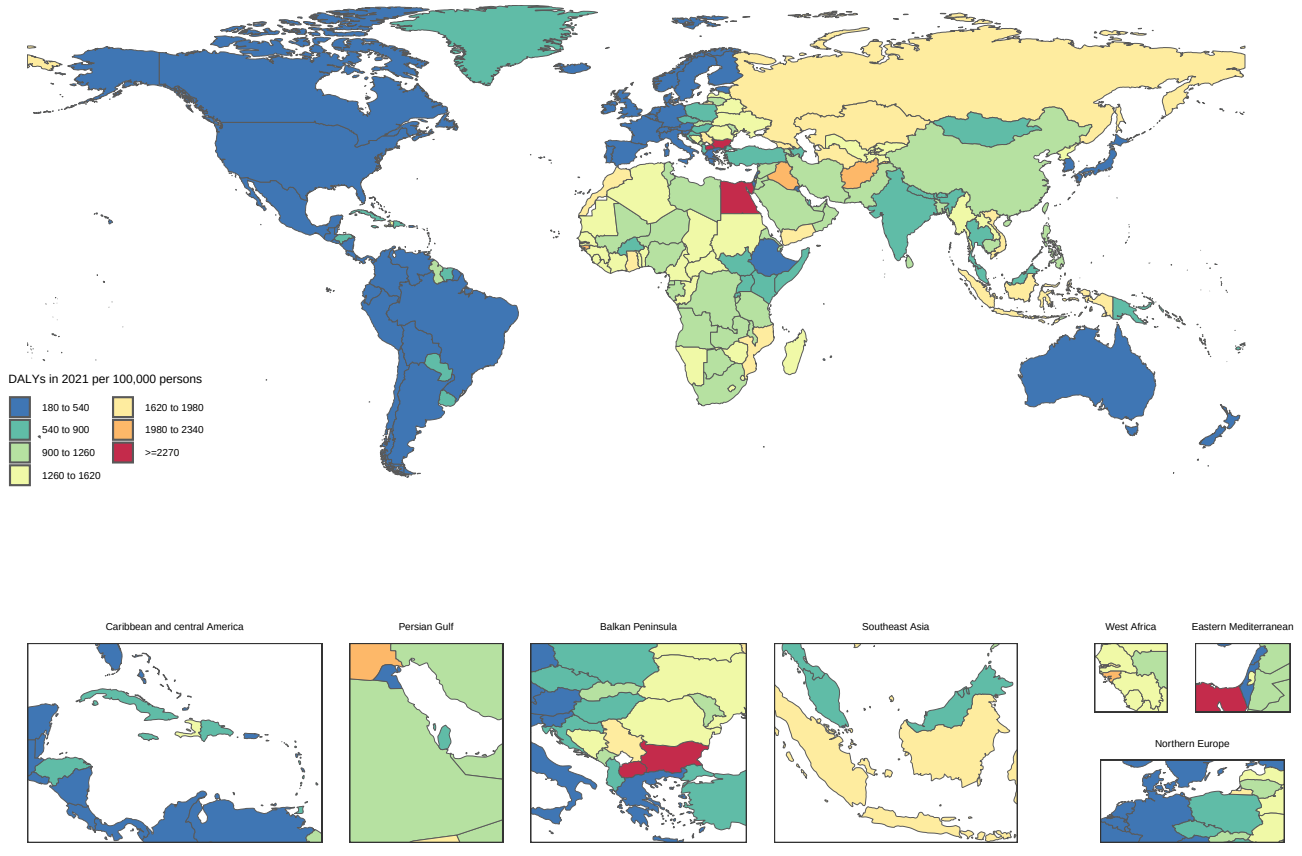

DALYs in 2021 per 100,000 persons
180 to 540
540 to 900
900 to 1260
1260 to 1620
1620 to 1980
1980 to 2340
>=2270
Caribbean and central America
Persian Gulf
Balkan Peninsula
Southeast Asia
West Africa
Eastern Mediterranean
Northern Europe

Figure S12: Global prevalence rate of ischemic stroke, by age and sex, in 2021; Red and blue shadows indicate 95% upper and lower uncertainty intervals, respectively. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

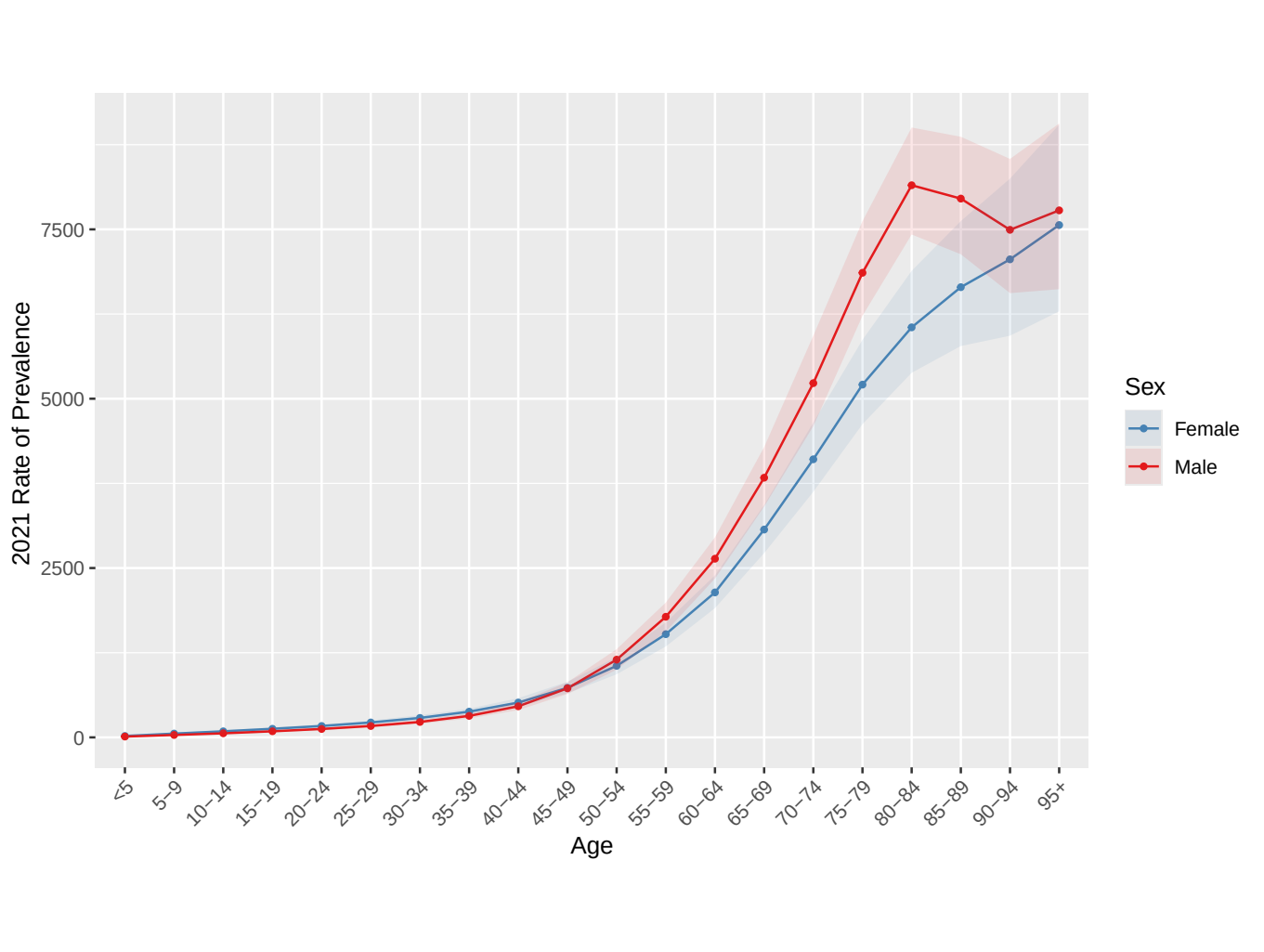

2021 Rate of Prevalence
7500
5000
2500
0
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age
Sex
Female
Male

Figure S13: Global number of prevalence of ischemic stroke per 100,000 population, by age and sex, in 2021. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

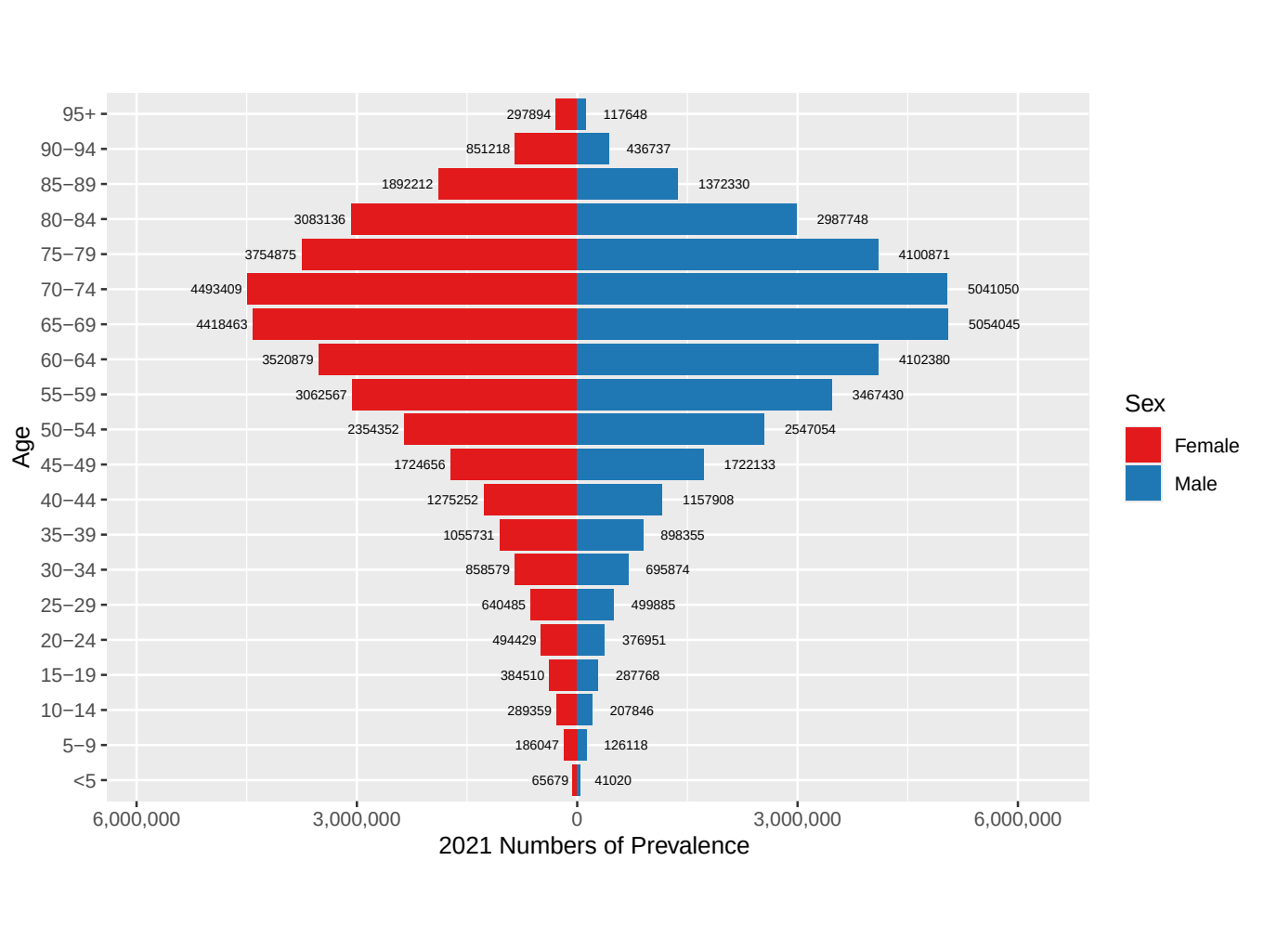

95+
90−94
85−89
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
5−9
<5
297894
117648
851218
436737
1892212
1372330
3083136
2987748
3754875
4100871
4493409
5041050
4418463
5054045
3520879
4102380
3062567
3467430
2354352
2547054
1724656
1722133
1275252
1157908
1055731
898355
858579
695874
640485
499885
494429
376951
384512
287768
289359
207846
186047
126118
65679
41020
Age
6,000,000
3,000,000
0
3,000,000
6,000,000
2021 Numbers of Prevalence
Sex
Female
Male

Figure S14: Number of incident cases globally and incidence of ischemic stroke, per 100,000 population by age and sex in 2021. Dotted and dashed lines indicate 95% upper and lower uncertainty intervals, respectively. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

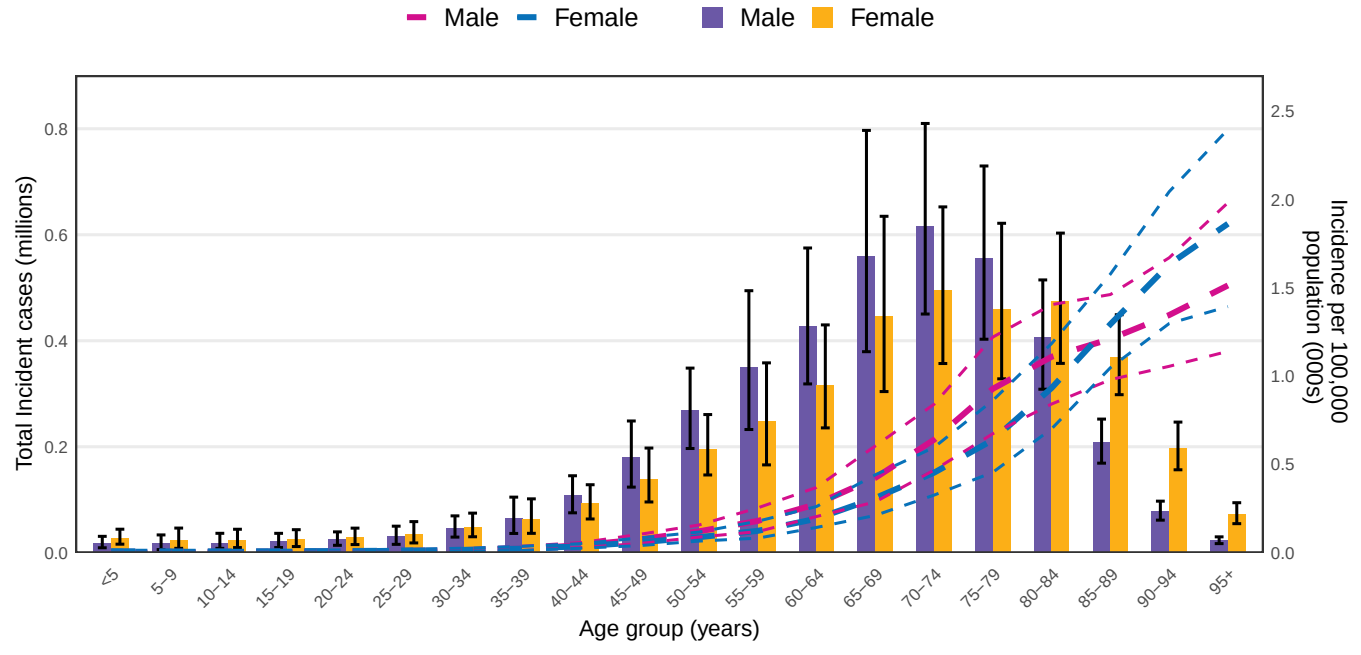

Male
Female
Male
Female
Total Incident cases (millions)
0.0
0.2
0.4
0.6
0.8
Incidence per 100,000 population (000s)
0.0
0.5
1.0
1.5
2.0
2.5
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age group (years)

Figure S15: Global incidence rate of ischemic stroke, per 100,000 population, by age and sex, in 2021; Red and blue shadows indicate 95% upper and lower uncertainty intervals, respectively. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

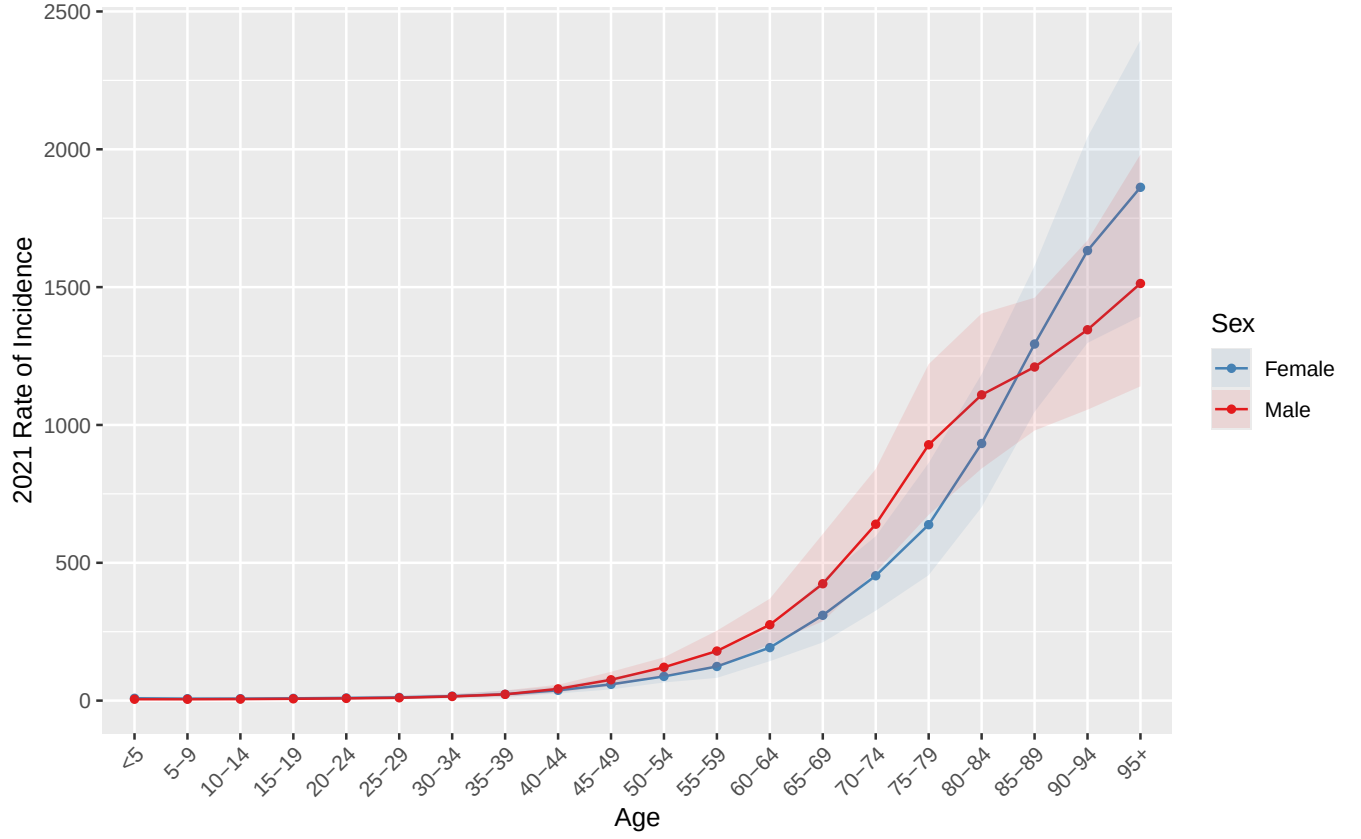

2500
2000
1500
1000
500
0
2021 Rate of Incidence
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age
Sex
Female
Male

Figure S16: Global number of incidence of ischemic stroke, per 100,000 population, by age and sex, in 2021. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

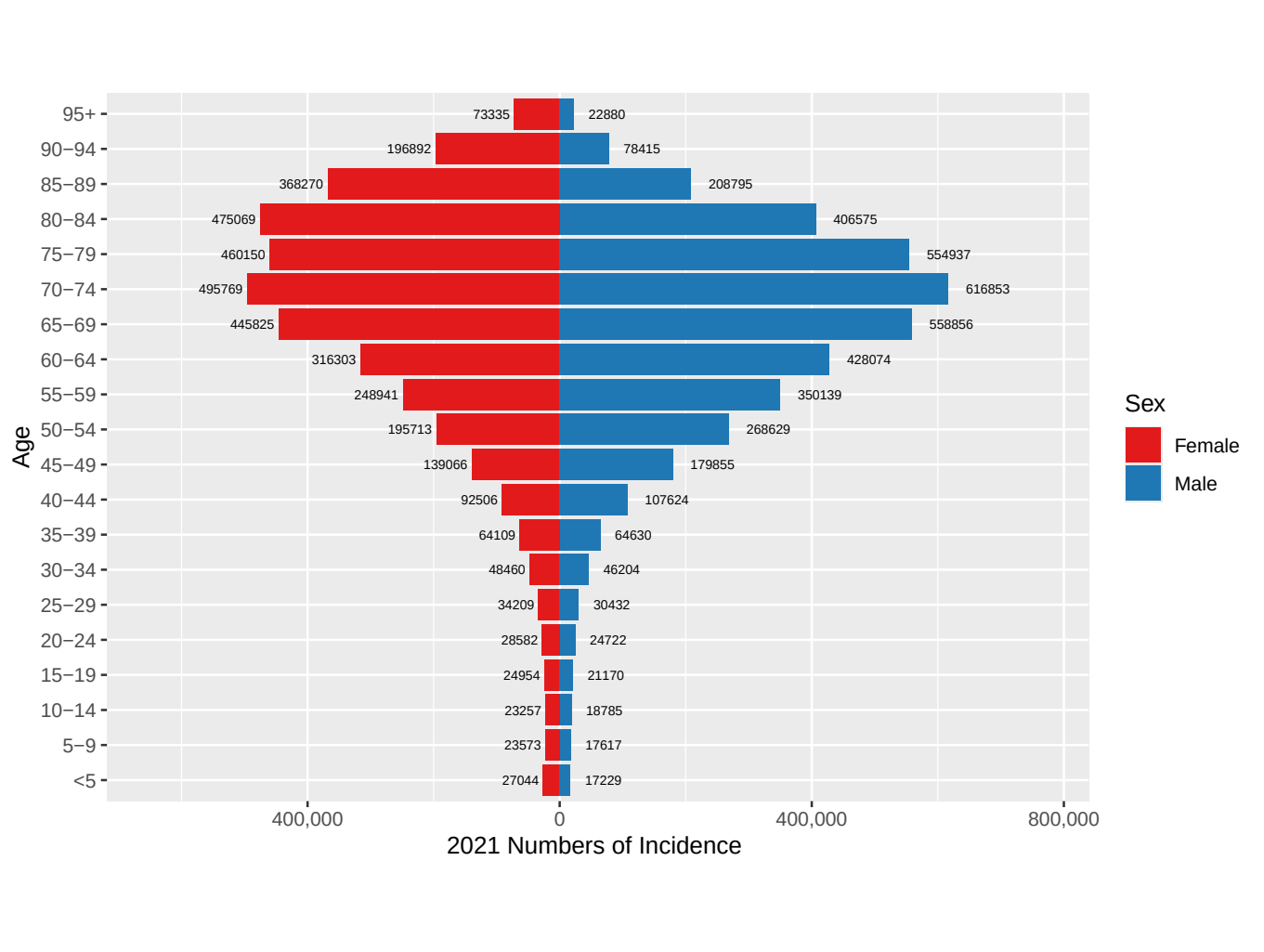

95+
90−94
85−89
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
5−9
<5
73335
22880
196892
78415
368270
208795
475069
406575
460150
554937
495769
616853
445825
558856
316303
428074
248941
350139
195713
268629
139066
179855
92506
107624
64109
64630
48460
46204
34209
30432
28582
24722
24954
21170
23257
18785
23573
17617
27044
17229
400,000
0
400,000
800,000
Age
2021 Numbers of Incidence
Sex
Female
Male

Figure S17: Number of death cases globally and death rate of ischemic stroke, per 100,000 population by age and sex in 2021. Dotted and dashed lines indicate 95% upper and lower uncertainty intervals, respectively. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

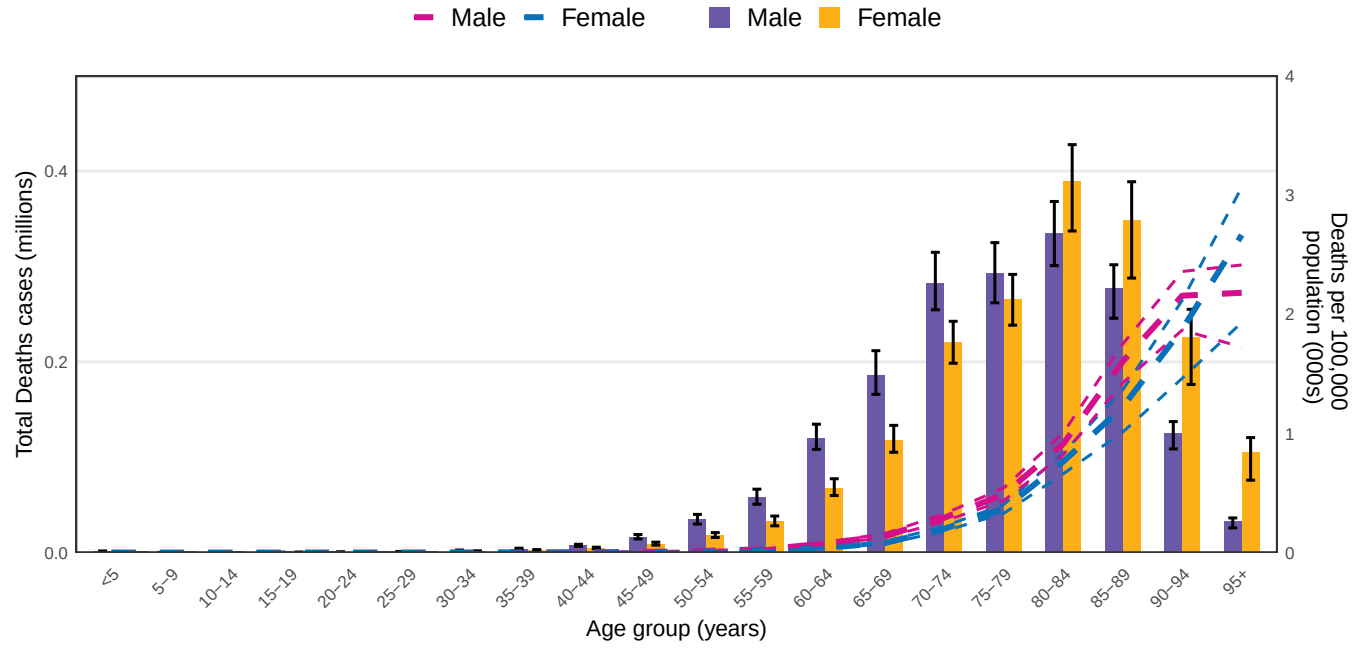

Male
Female
Male
Female
Total Deaths cases (millions)
0.0
0.2
0.4
Deaths per 100,000 population (000s)
0
1
2
3
4
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age group (years)

Figure S18: Global death rate of ischemic stroke, per 100,000 population, by age and sex, in 2021; Red and blue shadows indicate 95% upper and lower uncertainty intervals, respectively. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

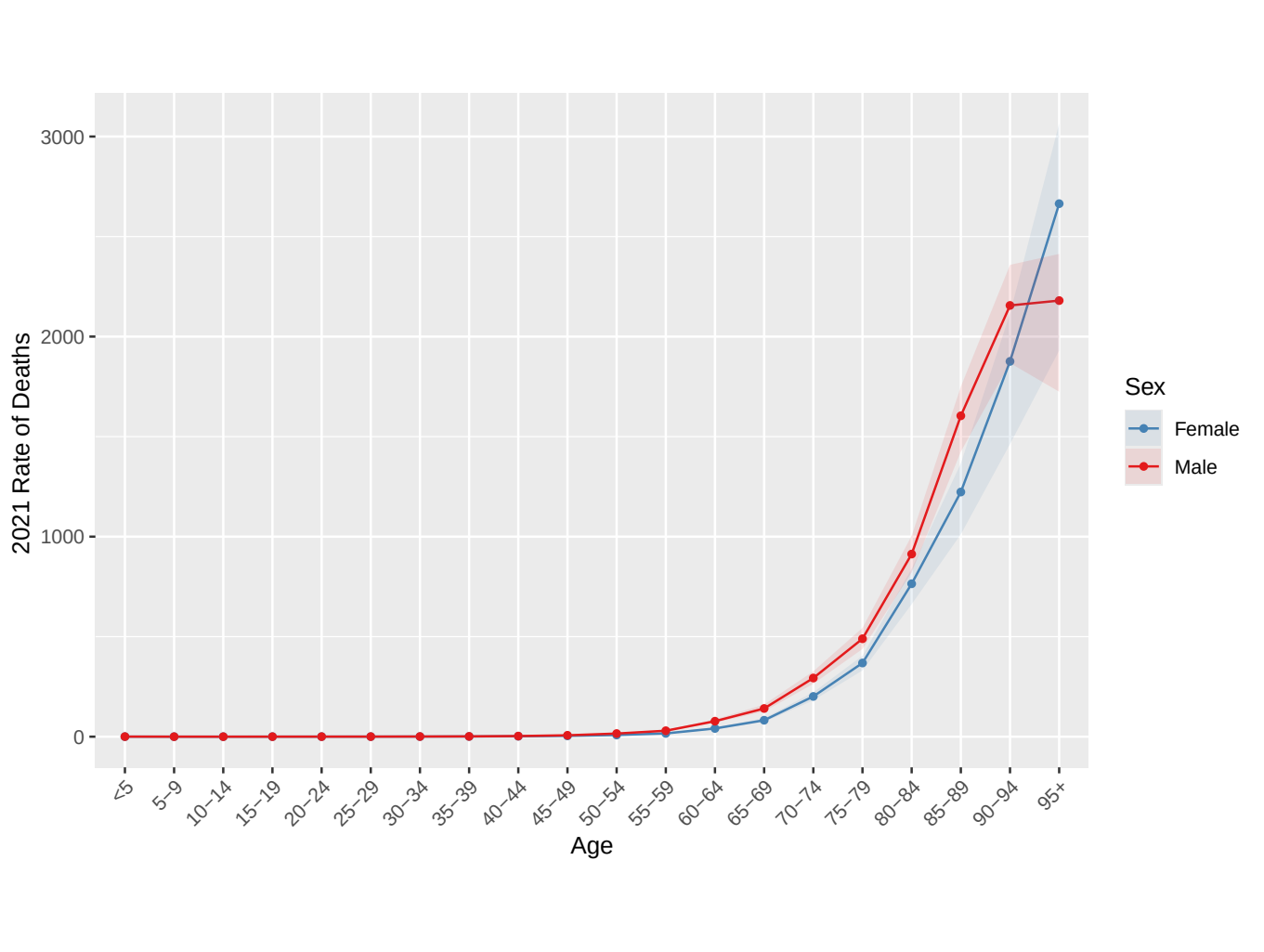

3000
2000
1000
0
2021 Rate of Deaths
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age
Sex
Female
Male

Figure S19: Global number of deaths of ischemic stroke, per 100,000 population, by age and sex, in 2021. (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

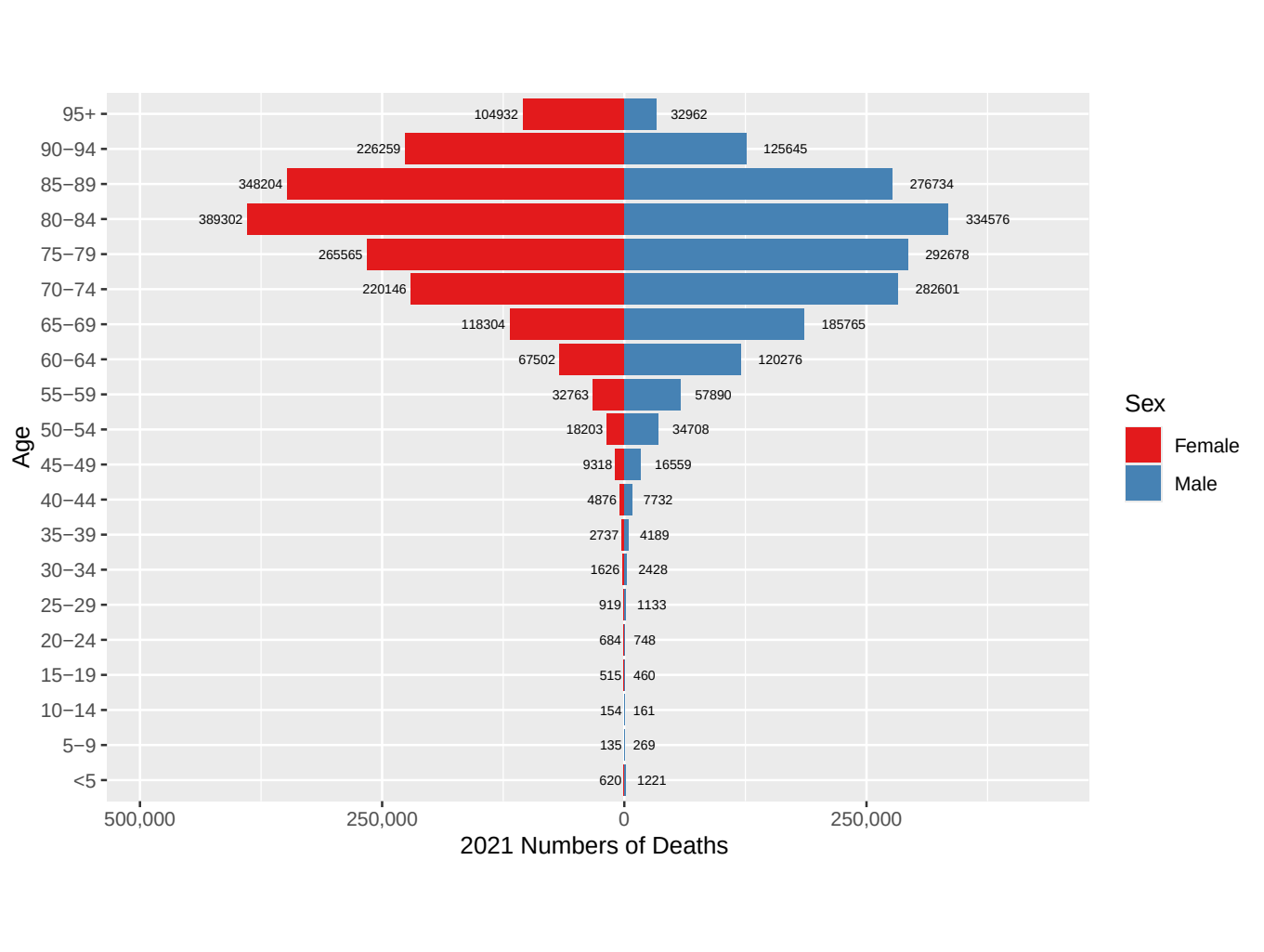

95+
90−94
85−89
80−84
75−79
70−74
65−69
60−64
55−59
50−54
45−49
40−44
35−39
30−34
25−29
20−24
15−19
10−14
5−9
<5
104932
32962
226259
125645
348204
276734
389302
334576
265565
292678
220146
282601
118304
185765
67502
120276
32763
57890
18203
34708
9318
16559
4876
7732
2737
4189
1626
2428
919
1133
684
748
515
460
154
161
135
269
620
1221
Age
500,000
250,000
0
250,000
2021 Numbers of Deaths
Sex
Female
Male

Figure S20: Number of DALYs globally and DALYs rate of ischemic stroke, per 100,000 population by age and sex in 2021. Dotted and dashed lines indicate 95% upper and lower uncertainty intervals, respectively.

DALYs = disability-adjusted life years (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

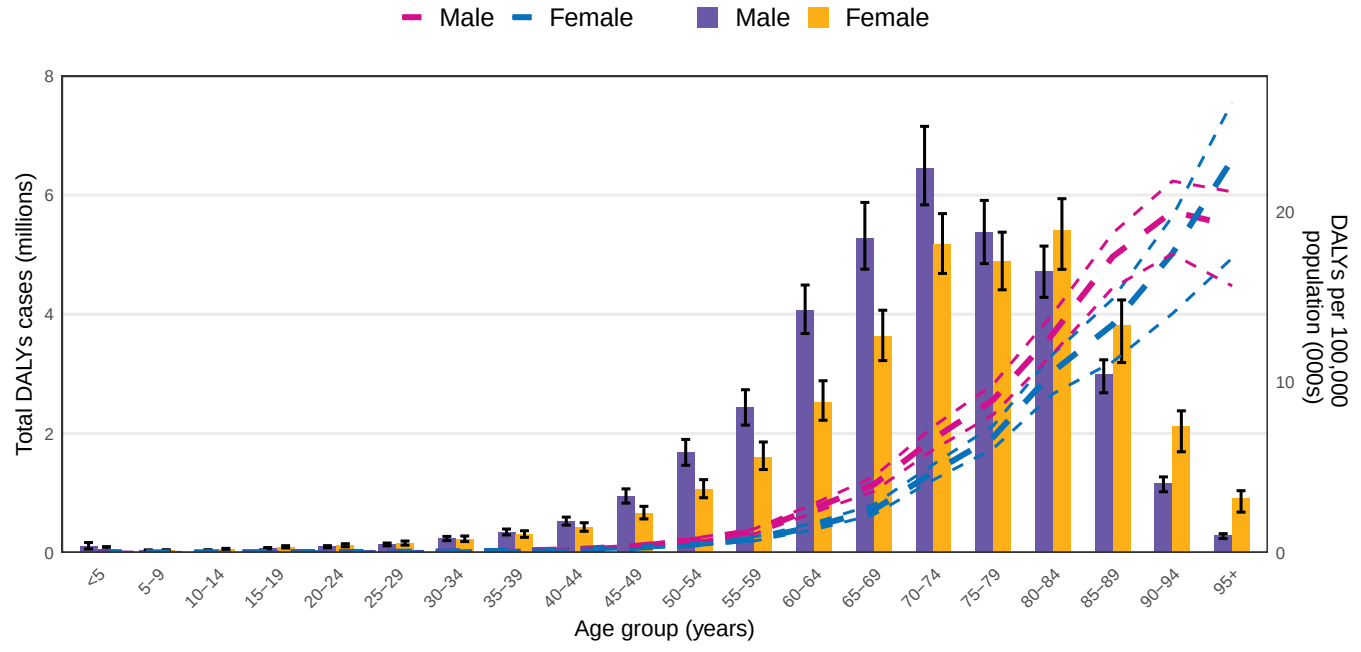

Male
Female
Male
Female
Total DALYs cases (millions)
DALYs per 100,000 population (000s)
0
2
4
6
8
0
10
20
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age group (years)

Figure S21: Global DALYs rate of ischemic stroke, per 100,000 population, by age and sex, in 2021; Red and blue shadows indicate 95% upper and lower uncertainty intervals, respectively.

DALYs = disability-adjusted life years (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

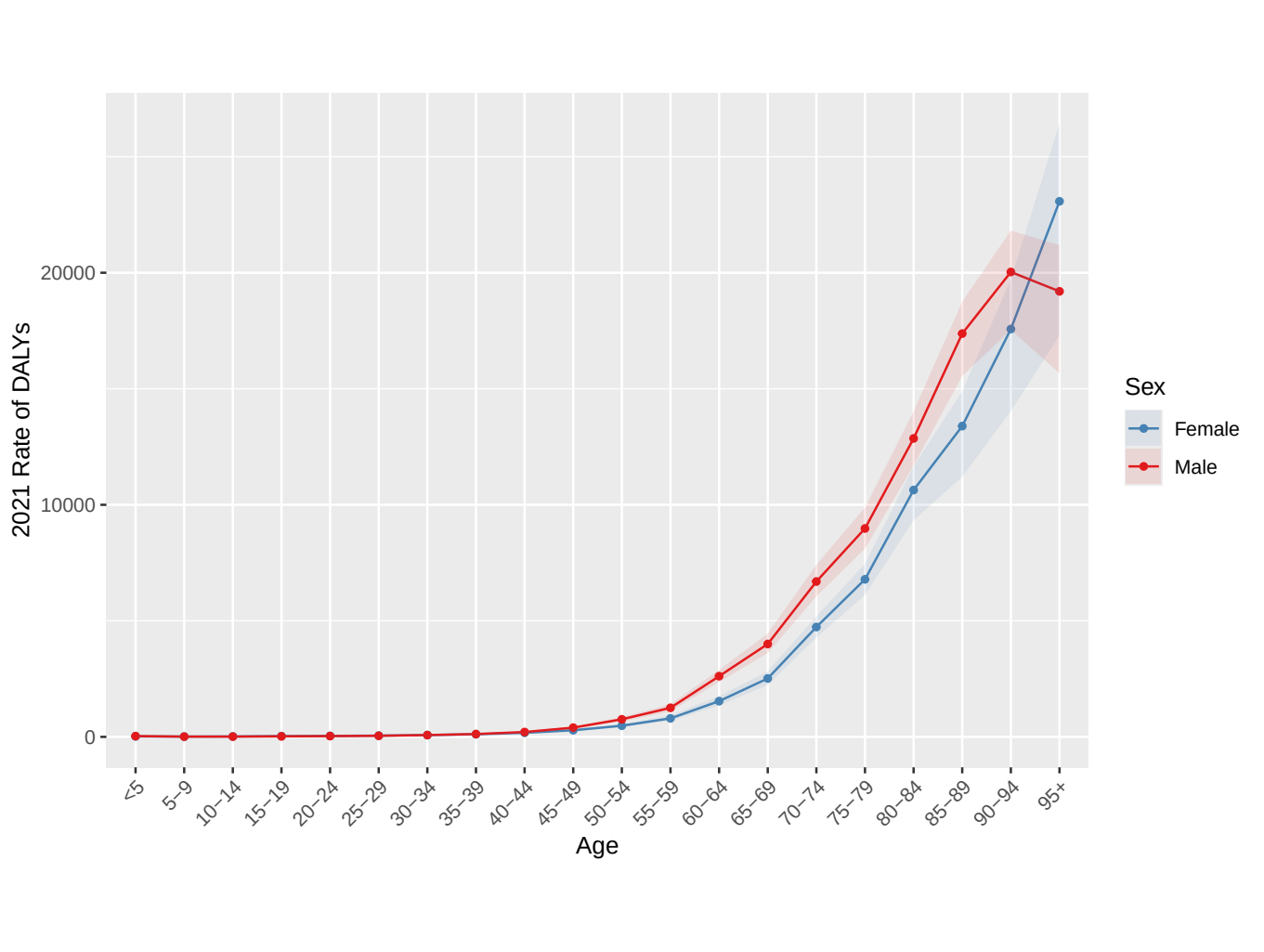

2021 Rate of DALYs
20000
10000
0
<5
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age
Sex
Female
Male

Figure S22: Global number of DALYs of ischemic stroke, per 100,000 population, by age and sex, in 2021.

DALYs = disability-adjusted life years (Generated from data available from http://ghdx.healthdata.org/gbd-results-tool).

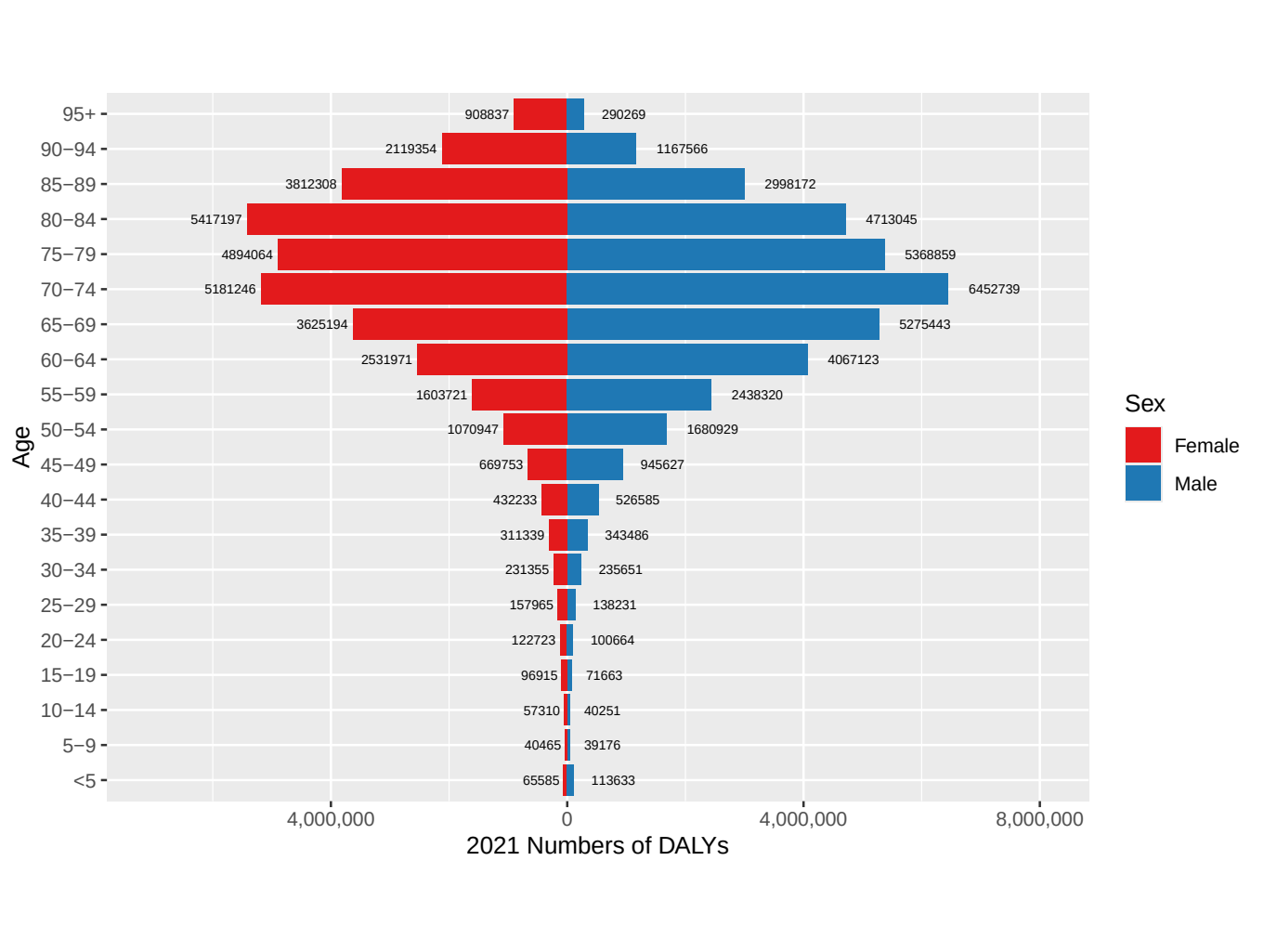

95+
90–94
85–89
80–84
75–79
70–74
65–69
60–64
55–59
50–54
45–49
40–44
35–39
30–34
25–29
20–24
15–19
10–14
5–9
<5
Age
908837
290269
2119354
1167566
3812308
2998172
5417197
4713045
4894064
5368859
5181246
6452739
3625194
5275443
2531971
4067123
1603721
2438320
1070947
1680929
669753
945627
432233
526585
311339
343486
231355
235651
157965
138231
122723
100664
96915
71663
57310
40251
40465
39176
65585
113633
4,000,000
0
4,000,000
8,000,000
2021 Numbers of DALYs
Sex
Female
Male

Figure S23: Disability-adjusted life years (DALYs) rates of ischemic stroke for the 22 Global Burden of Disease regions by sociodemographic index, 1990–2021. Thirty-two points are plotted for each region and show the observed age-standardized DALY rates from 1990 to 2021 for that region.

DALYs = disability-adjusted life years. Expected values, based on sociodemographic index and disease rates in all locations, are shown as a solid line. Regions above the solid line represent a higher-than-expected burden and regions below the line show a lower-than-expected burden (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

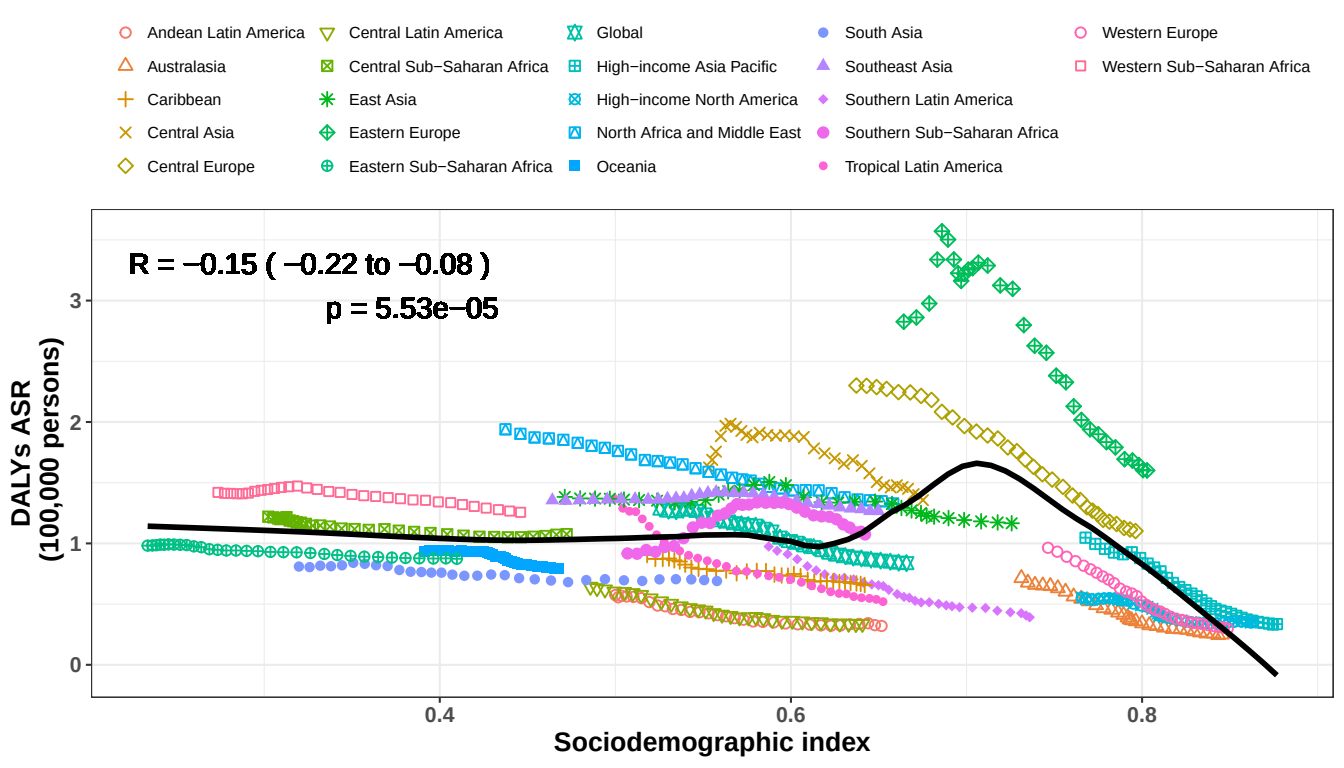

Andean Latin America
Australasia
Caribbean
Central Asia
Central Europe
Central Latin America
Central Sub−Saharan Africa
East Asia
Eastern Europe
Eastern Sub−Saharan Africa
Global
High−income Asia Pacific
High−income North America
North Africa and Middle East
Oceania
South Asia
Southeast Asia
Southern Latin America
Southern Sub−Saharan Africa
Tropical Latin America
Western Europe
Western Sub−Saharan Africa
R = −0.15 (−0.22 to −0.08)
p = 5.53e−05
DALYs ASR
(100,000 persons)
0
1
2
3
0.4
0.6
0.8
Sociodemographic index

Figure S24: Percentage of disability-adjusted life years (DALYs) due to ischemic stroke attributable to each risk factor for the 21 Global Burden of Disease regions in 2021. (Top 1–6)

DALYs = disability-adjusted life years. (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

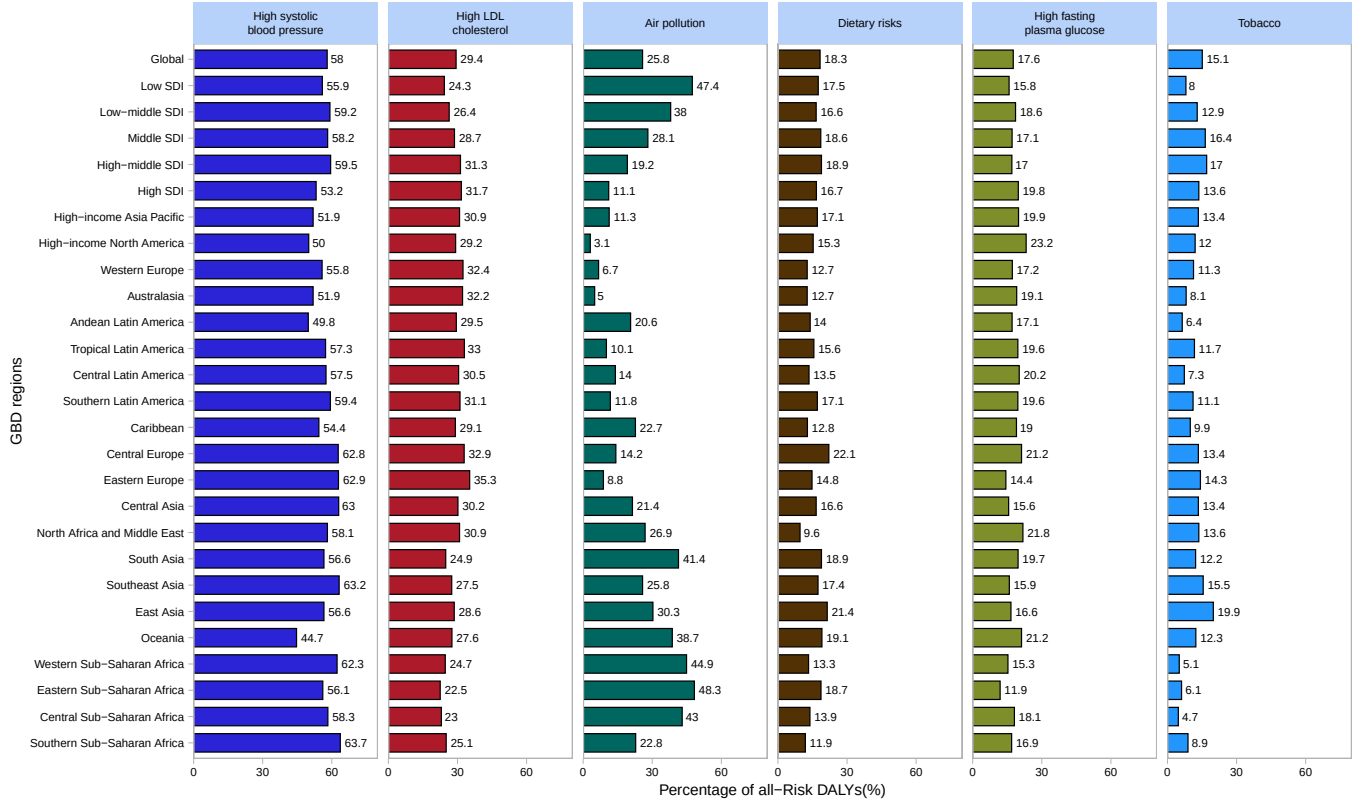

High systolic blood pressure
High LDL cholesterol
Air pollution
Dietary risks
High fasting plasma glucose
Tobacco
GBD regions
Global 58 29.4 25.8 18.3 17.6 15.1
Low SDI 55.9 24.3 47.4 17.5 15.8 8
Low−middle SDI 59.2 26.4 38 16.6 18.6 12.9
Middle SDI 58.2 28.7 28.1 18.6 17.1 16.4
High−middle SDI 59.5 31.3 19.2 18.9 17 17
High SDI 53.2 31.7 11.1 16.7 19.8 13.6
High−income Asia Pacific 51.9 30.9 11.3 17.1 19.9 13.4
High−income North America 50 29.2 3.1 15.3 23.2 12
Western Europe 55.8 32.4 6.7 12.7 17.2 11.3
Australasia 51.9 32.2 5 12.7 19.1 8.1
Andean Latin America 49.8 29.5 20.6 14 17.1 6.4
Tropical Latin America 57.3 33 10.1 15.6 19.6 11.7
Central Latin America 57.5 30.5 14 13.5 20.2 7.3
Southern Latin America 59.4 31.1 11.8 17.1 19.6 11.1
Caribbean 54.4 29.1 22.7 12.8 19 9.9
Central Europe 62.8 32.9 14.2 22.1 21.2 13.4
Eastern Europe 62.9 35.3 8.8 14.8 14.4 14.3
Central Asia 63 30.2 21.4 16.6 15.6 13.4
North Africa and Middle East 58.1 30.9 26.9 9.6 21.8 13.6
South Asia 56.6 24.9 41.4 18.9 19.7 12.2
Southeast Asia 63.2 27.5 25.8 17.4 15.9 15.5
East Asia 56.6 28.6 30.3 21.4 16.6 19.9
Oceania 44.7 27.6 38.7 19.1 21.2 12.3
Western Sub−Saharan Africa 62.3 24.7 44.9 13.3 15.3 5.1
Eastern Sub−Saharan Africa 56.1 22.5 48.3 18.7 11.9 6.1
Central Sub−Saharan Africa 58.3 23 43 13.9 18.1 4.7
Southern Sub−Saharan Africa 63.7 25.1 22.8 11.9 16.9 8.9
0 30 60
Percentage of all−Risk DALYs(%)

Figure S25: Percentage of disability-adjusted life years (DALYs) due to ischemic stroke attributable to each risk factor for the 21 Global Burden of Disease regions in 2021. (Top 7–12)

DALYs = disability-adjusted life years, GBD = Global Burden of Disease. (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

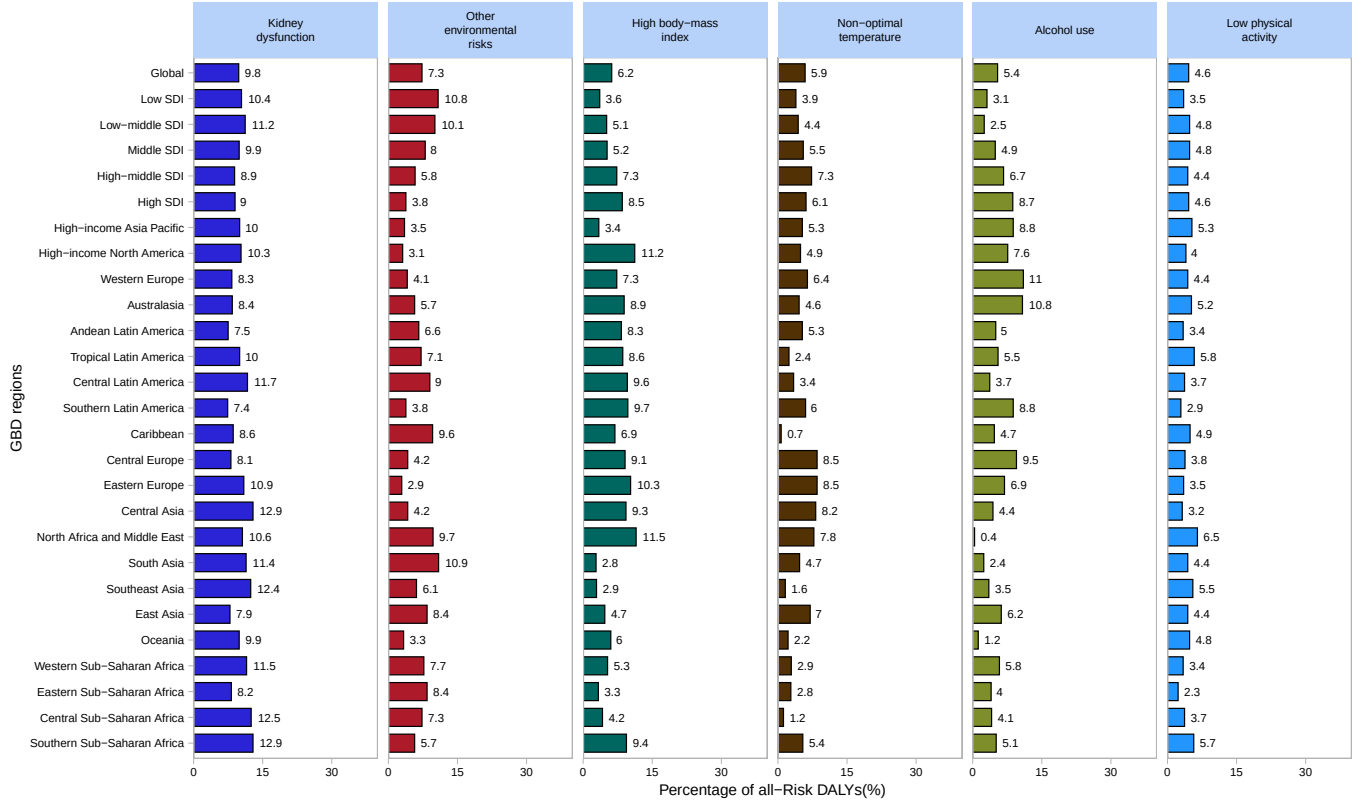

Kidney dysfunction
Other environmental risks
High body−mass index
Non−optimal temperature
Alcohol use
Low physical activity
GBD regions
Global
Low SDI
Low−middle SDI
Middle SDI
High−middle SDI
High SDI
High−income Asia Pacific
High−income North America
Western Europe
Australasia
Andean Latin America
Tropical Latin America
Central Latin America
Southern Latin America
Caribbean
Central Europe
Eastern Europe
Central Asia
North Africa and Middle East
South Asia
Southeast Asia
East Asia
Oceania
Western Sub−Saharan Africa
Eastern Sub−Saharan Africa
Central Sub−Saharan Africa
Southern Sub−Saharan Africa
9.8 7.3 6.2 5.9 5.4 4.6
10.4 10.8 3.6 3.9 3.1 3.5
11.2 10.1 5.1 4.4 2.5 4.8
9.9 8 5.2 5.5 4.9 4.8
8.9 5.8 7.3 7.3 6.7 4.4
9 3.8 8.5 6.1 8.7 4.6
10 3.5 3.4 5.3 8.8 5.3
10.3 3.1 11.2 4.9 7.6 4
8.3 4.1 7.3 6.4 11 4.4
8.4 5.7 8.9 4.6 10.8 5.2
7.5 6.6 8.3 5.3 5 3.4
10 7.1 8.6 2.4 5.5 5.8
11.7 9 9.6 3.4 3.7 3.7
7.4 3.8 9.7 6 8.8 2.9
8.6 9.6 6.9 0.7 4.7 4.9
8.1 4.2 9.1 8.5 9.5 3.8
10.9 2.9 10.3 8.5 6.9 3.5
12.9 4.2 9.3 8.2 4.4 3.2
10.6 9.7 11.5 7.8 0.4 6.5
11.4 10.9 2.8 4.7 2.4 4.4
12.4 6.1 2.9 1.6 3.5 5.5
7.9 8.4 4.7 7 6.2 4.4
9.9 3.3 6 2.2 1.2 4.8
11.5 7.7 5.3 2.9 5.8 3.4
8.2 8.4 3.3 2.8 4 2.3
12.5 7.3 4.2 1.2 4.1 3.7
12.9 5.7 9.4 5.4 5.1 5.7
0 15 30
Percentage of all−Risk DALYs(%)

Figure S26: Percentage of disability-adjusted life years (DALYs) due to ischemic stroke attributable to each risk factor by age groups in 2021. (Top 1–6)

DALYs = disability-adjusted life years, LDL = Low-Density Lipoprotein. (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

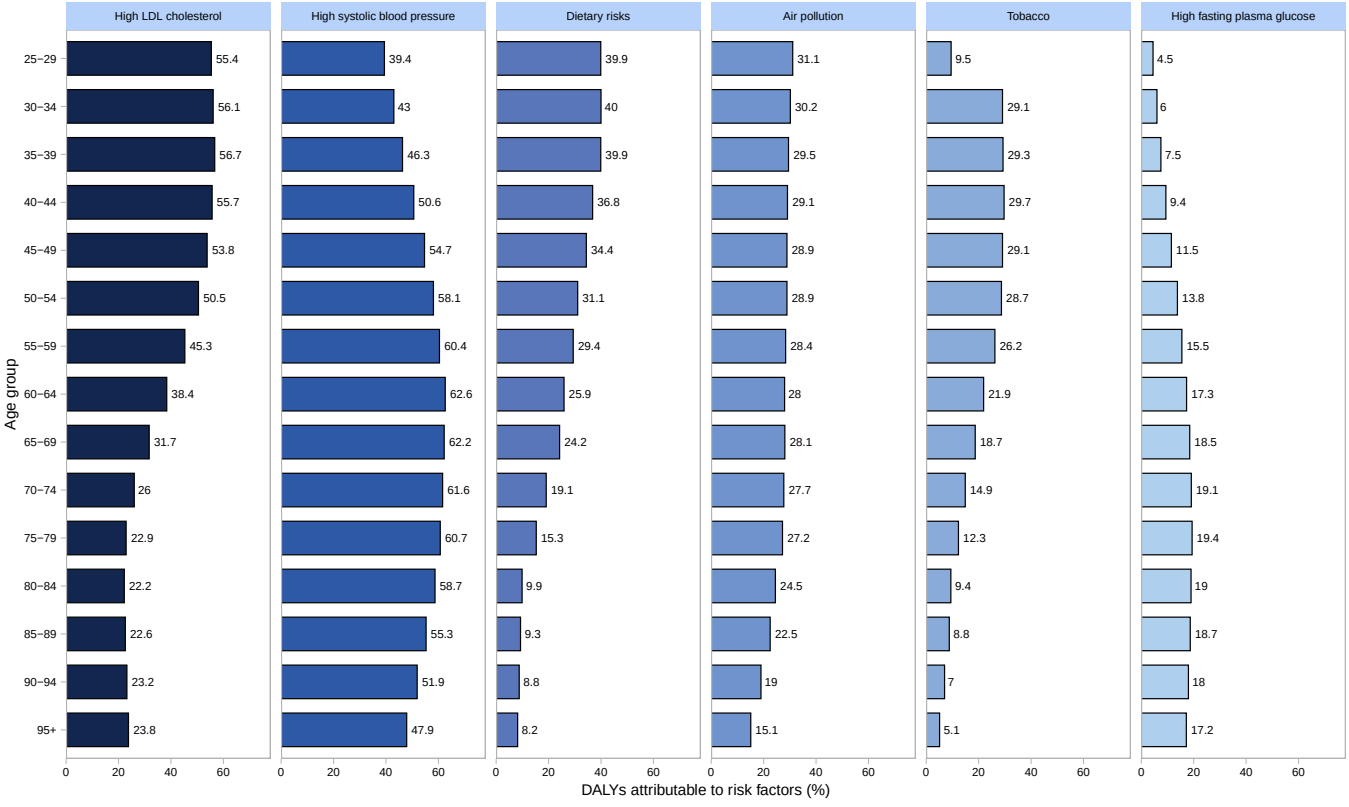

High LDL cholesterol
High systolic blood pressure
Dietary risks
Air pollution
Tobacco
High fasting plasma glucose
Age group
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
55.4
56.1
56.7
55.7
53.8
50.5
45.3
38.4
31.7
26
22.9
22.2
22.6
23.2
23.8
39.4
43
46.3
50.6
54.7
58.1
60.4
62.6
62.2
61.6
60.7
58.7
55.3
51.9
47.9
39.9
40
39.9
36.8
34.4
31.1
29.4
25.9
24.2
19.1
15.3
9.9
9.3
8.8
8.2
31.1
30.2
29.5
29.1
28.9
28.9
28.4
28
28.1
27.7
27.2
24.5
22.5
19
15.1
9.5
29.1
29.3
29.7
29.1
28.7
26.2
21.9
18.7
14.9
12.3
9.4
8.8
7
5.1
4.5
6
7.5
9.4
11.5
13.8
15.5
17.3
18.5
19.1
19.4
19
18.7
18
17.2
0
20
40
60
DALYs attributable to risk factors (%)

Figure S27: Percentage of disability-adjusted life years (DALYs) due to ischemic stroke attributable to each risk factor by age groups in 2021. (Top 7–12)

DALYs = disability-adjusted life years. (Generated from data available at https://ghdx.healthdata.org/gbd-results-tool).

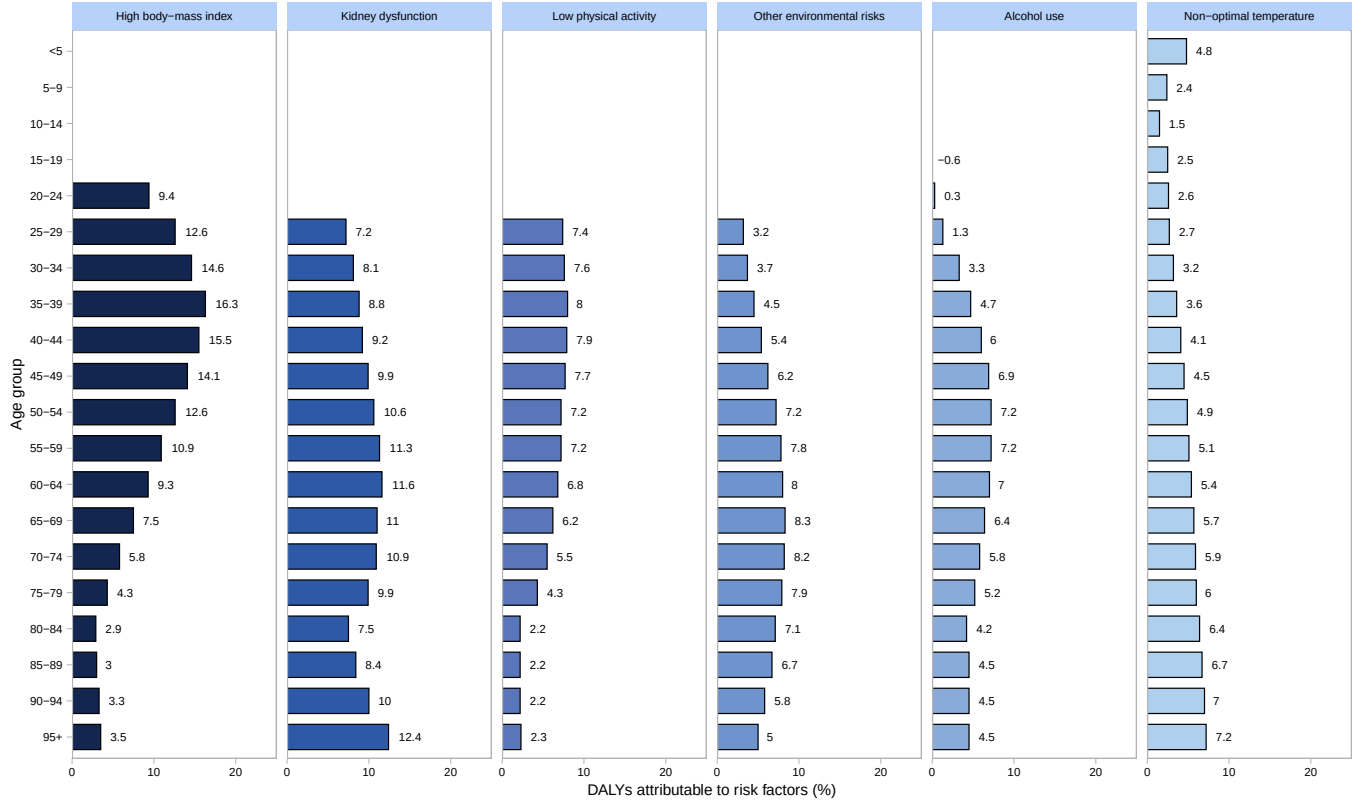

High body−mass index
Kidney dysfunction
Low physical activity
Other environmental risks
Alcohol use
Non−optimal temperature
Age group
<5
5−9
10−14
15−19
20−24
25−29
30−34
35−39
40−44
45−49
50−54
55−59
60−64
65−69
70−74
75−79
80−84
85−89
90−94
95+
9.4
12.6
14.6
16.3
15.5
14.1
12.6
10.9
9.3
7.5
5.8
4.3
2.9
3
3.3
3.5
7.2
8.1
8.8
9.2
9.9
10.6
11.3
11.6
11
10.9
9.9
7.5
8.4
10
12.4
7.4
7.6
8
7.9
7.7
7.2
7.2
6.8
6.2
5.5
4.3
2.2
2.2
2.2
2.3
3.2
3.7
4.5
5.4
6.2
7.2
7.8
8
8.3
8.2
7.9
7.1
6.7
5.8
5
−0.6
0.3
1.3
3.3
4.7
6
6.9
7.2
7.2
7
6.4
5.8
5.2
4.2
4.5
4.5
4.5
4.8
2.4
1.5
2.5
2.6
2.7
3.2
3.6
4.1
4.5
4.9
5.1
5.4
5.7
5.9
6
6.4
6.7
7
7.2
0
10
20
DALYs attributable to risk factors (%)

Figure S28: A frontier analysis based on SDI and ischemic stroke DALYs rate from 1990 to 2021.

DALYs = disability-adjusted life years. SDI = sociodemographic index.

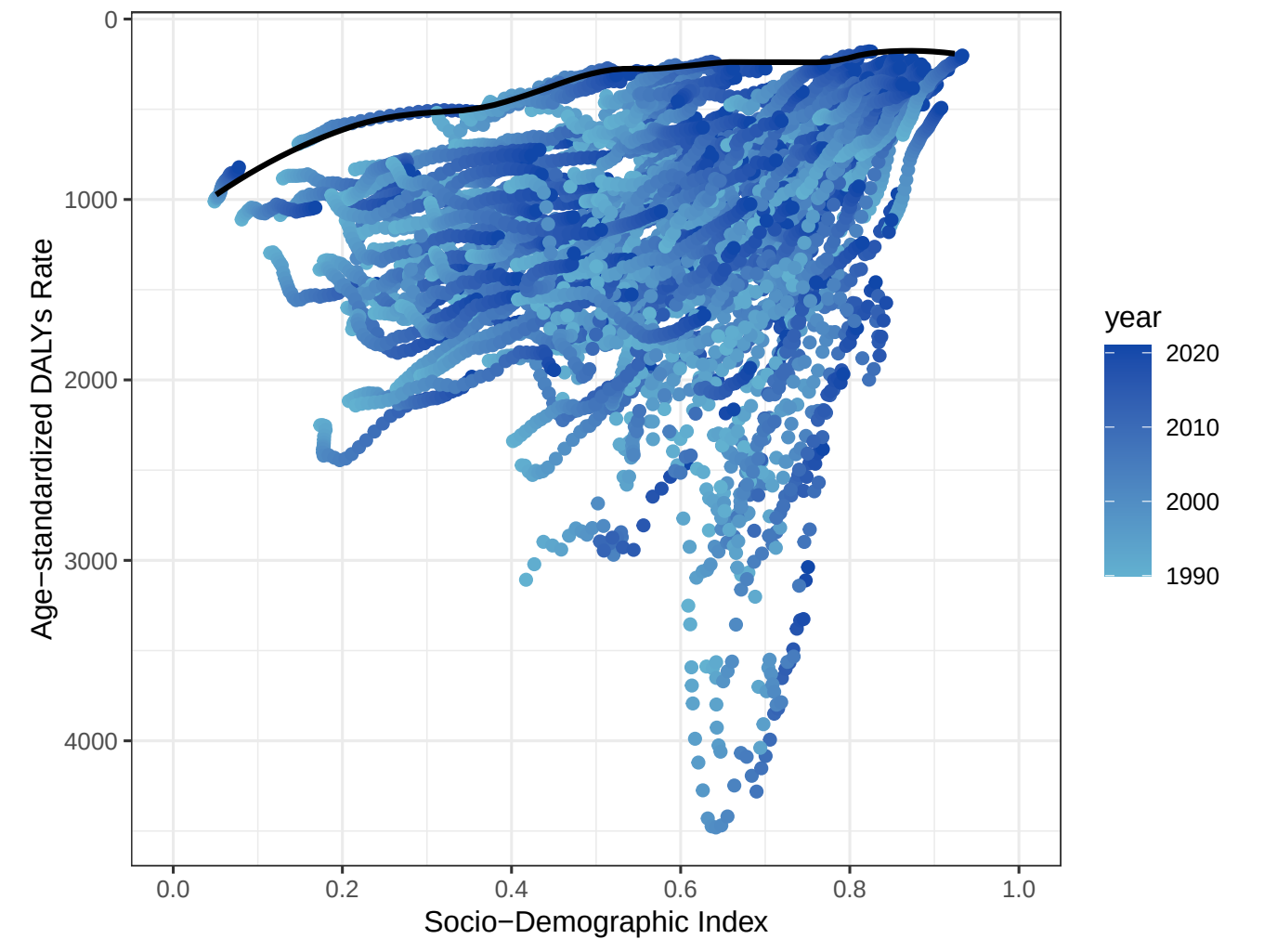

0
1000
2000
3000
4000
Age−standardized DALYs Rate
0.0
0.2
0.4
0.6
0.8
1.0
Socio−Demographic Index
year
2020
2010
2000
1990

Figure S29: Projected number of ASPR cases for ischemic stroke in males in 2036 according to the SDI.

SDI = sociodemographic index. ASPR = Age-Standardized Prevalence Rate.

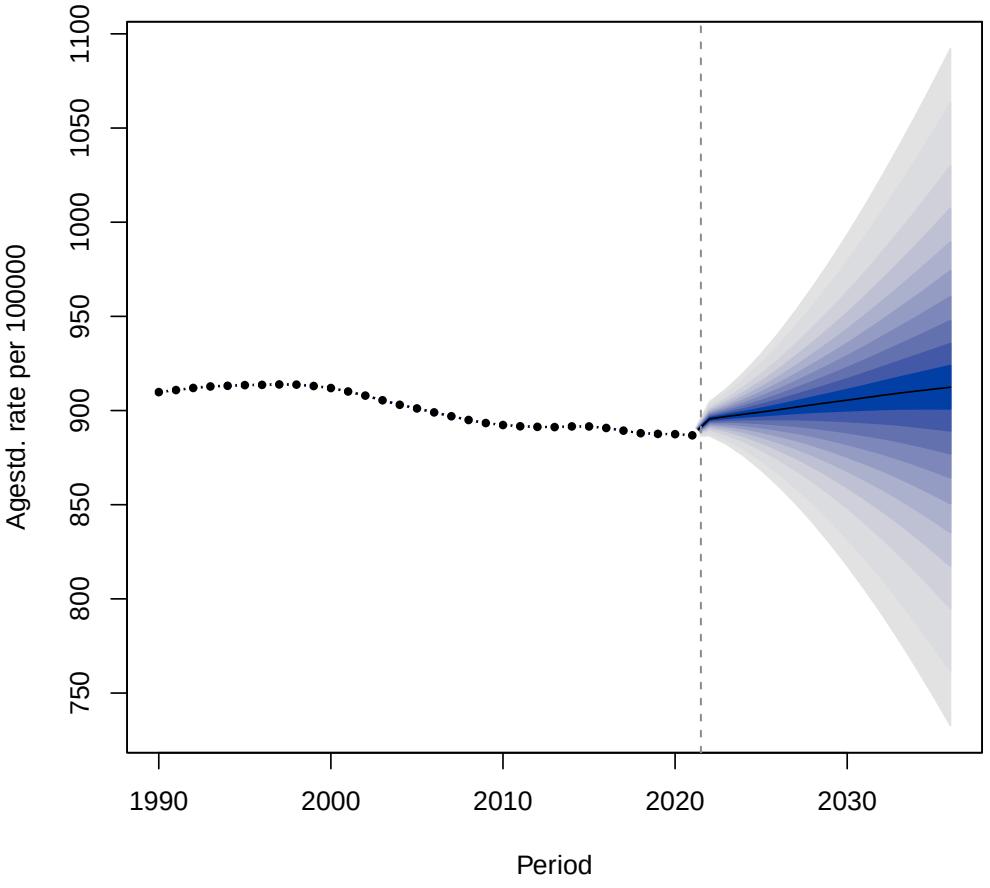

Agestd. rate per 100000
1100
1050
1000
950
900
850
800
750
1990
2000
2010
2020
2030
Period

Figure S30: Projected number of ASIR cases for ischemic stroke in males in 2036 according to the SDI.

SDI = sociodemographic index. ASIR = Age-Standardized Incidence Rate.

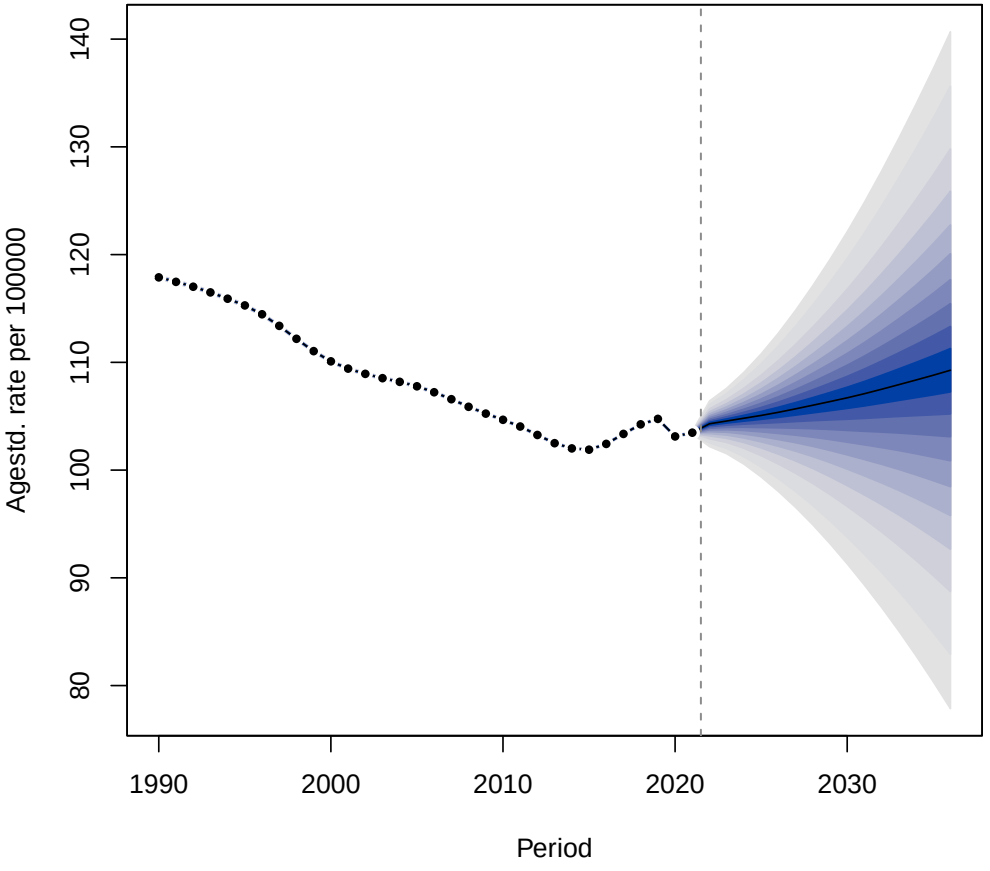

Agestd. rate per 100000
80
90
100
110
120
130
140
1990
2000
2010
2020
2030
Period

Figure S31: Projected number of ASDR cases for ischemic stroke in males in 2036 according to the SDI.

SDI = sociodemographic index. ASDR = Age-Standardized Death Rate.

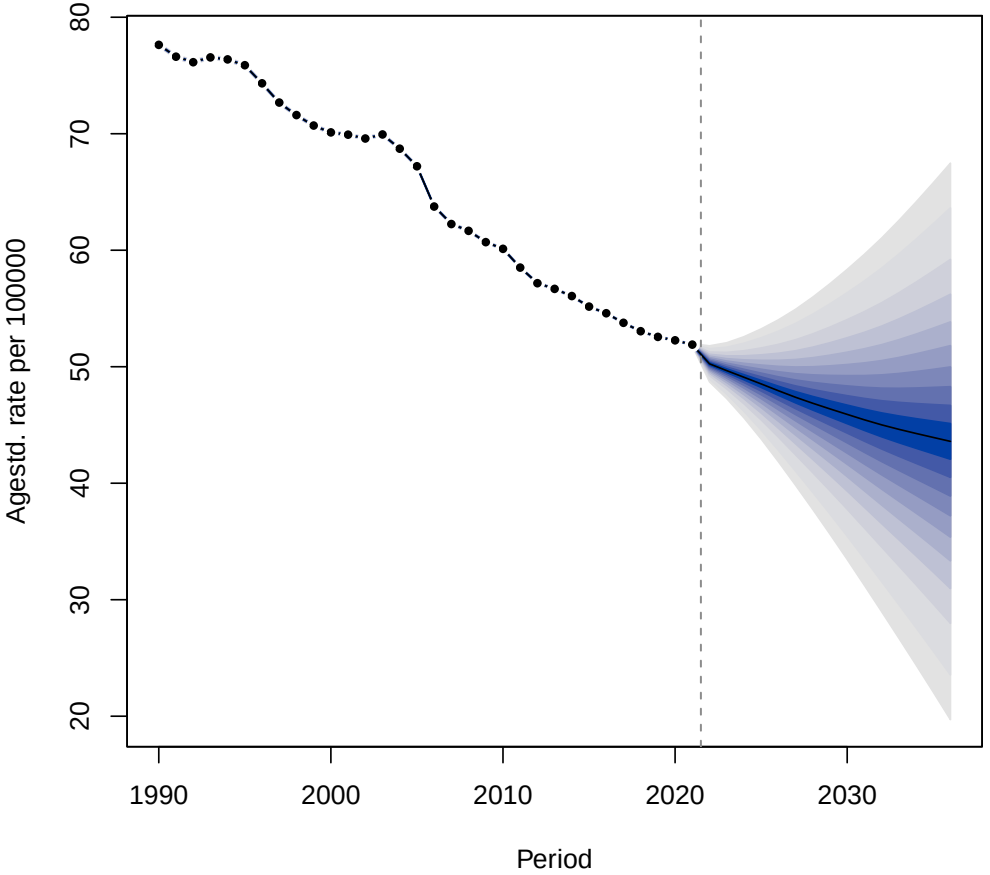

Agestd. rate per 100000
80
70
60
50
40
30
20
1990
2000
2010
2020
2030
Period

Figure S32: Projected number of new age-standardized DALY cases for ischemic stroke in males in 2036 according to the SDI.

DALYs = disability-adjusted life years. SDI = sociodemographic index.

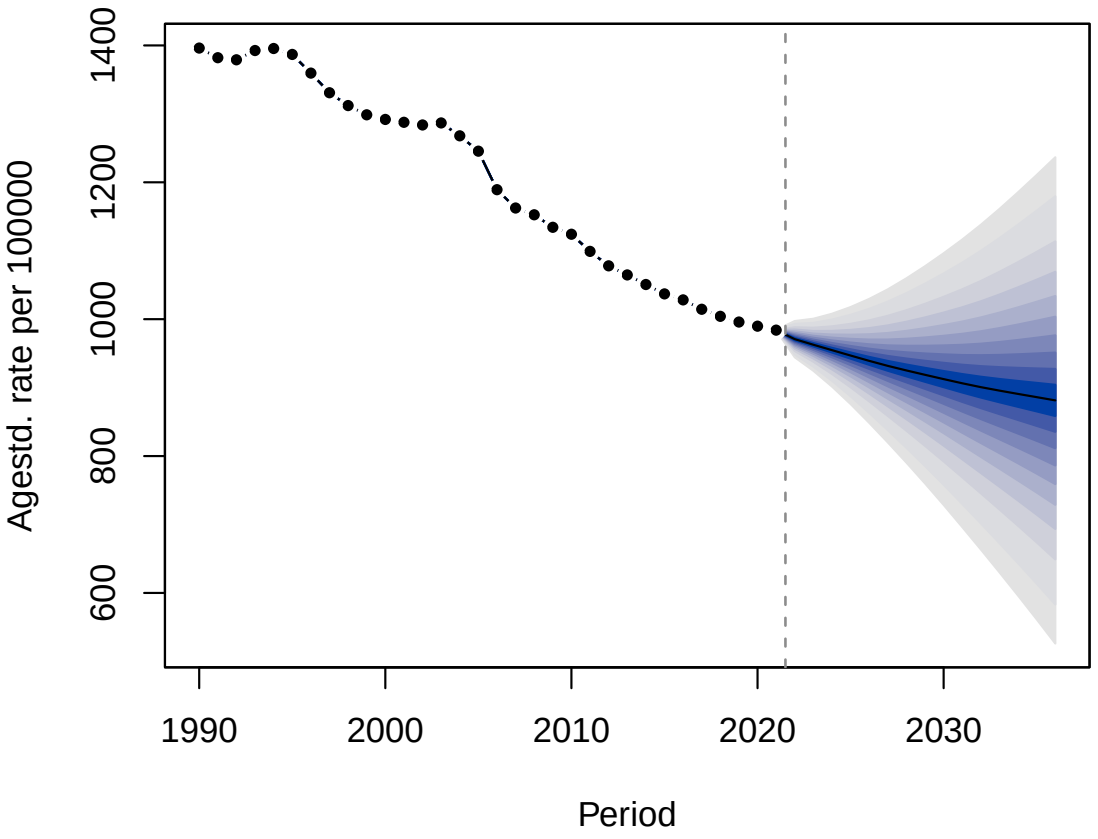

Agestd. rate per 100000
600
800
1000
1200
1400
1990
2000
2010
2020
2030
Period

Figure S33: Projected number of ASPR cases for ischemic stroke in females in 2036 according to the SDI.

SDI = sociodemographic index. ASPR = Age-Standardized Prevalence Rate.

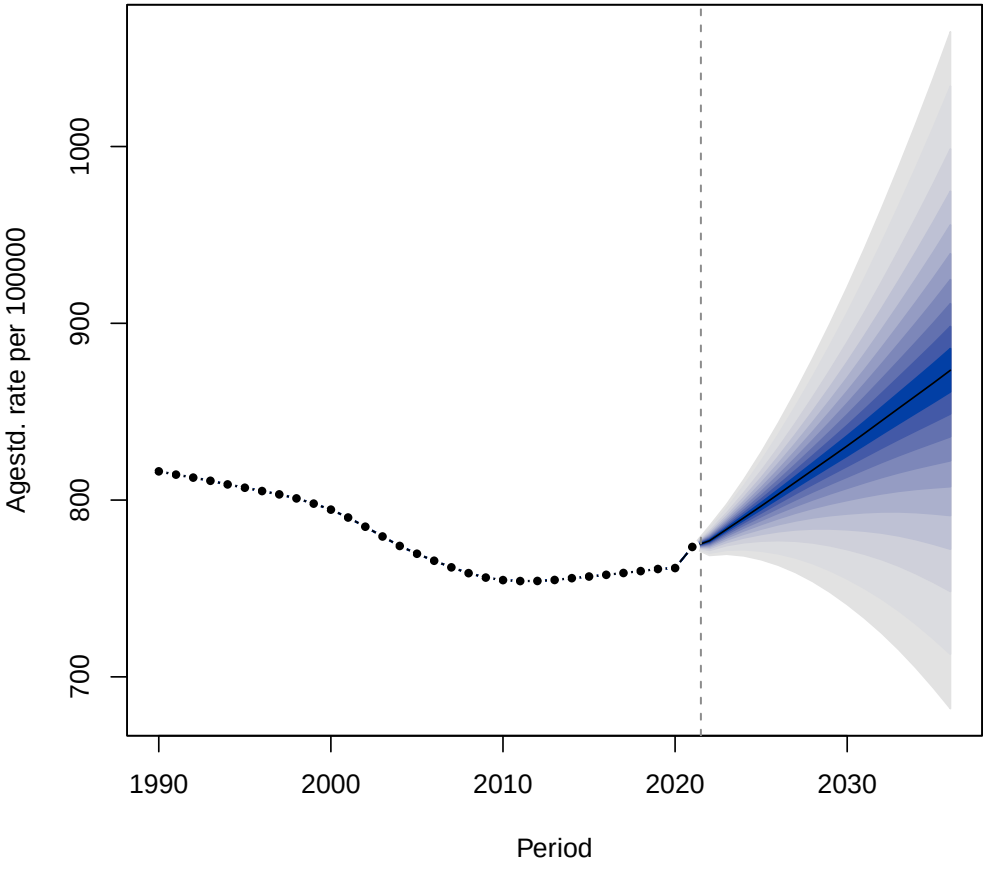

Agestd. rate per 100000
700
800
900
1000
1990
2000
2010
2020
2030
Period

Figure S34: Projected number of ASIR cases for ischemic stroke in females in 2036 according to the SDI.

SDI = sociodemographic index. ASIR = Age-Standardized Incidence Rate.

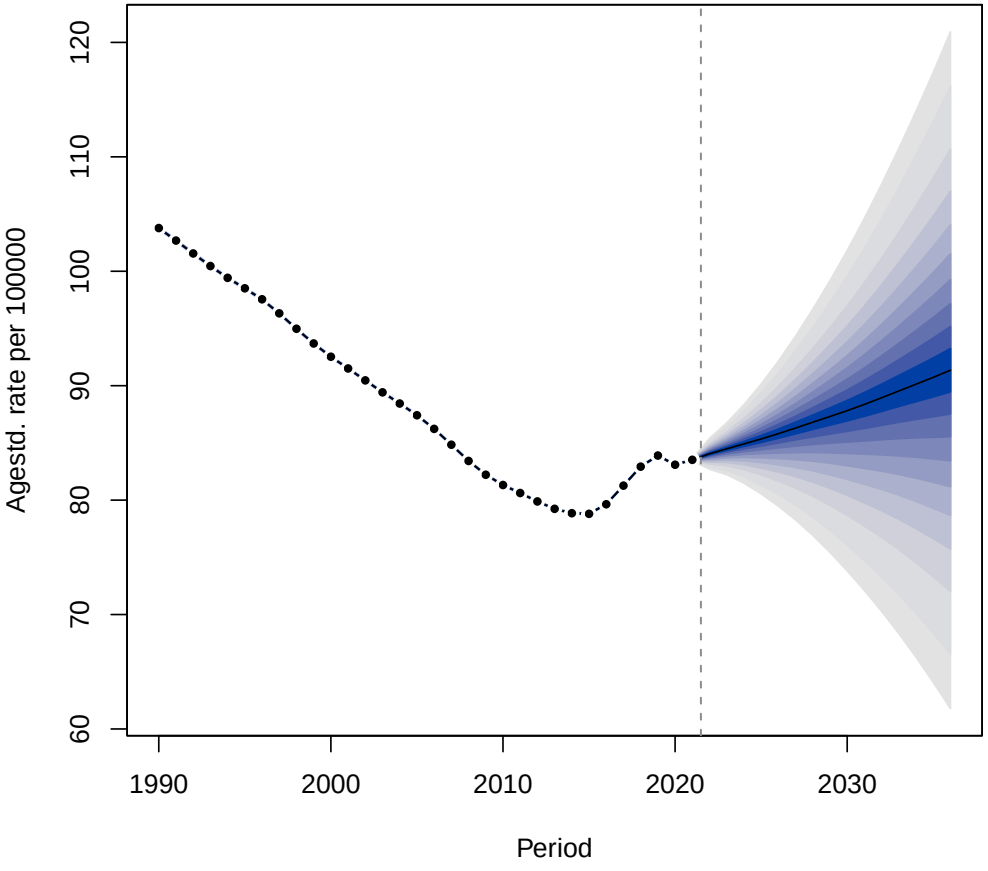

Agestd. rate per 100000
120
110
100
90
80
70
60
1990
2000
2010
2020
2030
Period

Figure S35: Projected number of ASDR cases for ischemic stroke in females in 2036 according to the SDI.

SDI = sociodemographic index. ASDR = Age-Standardized Death Rate.

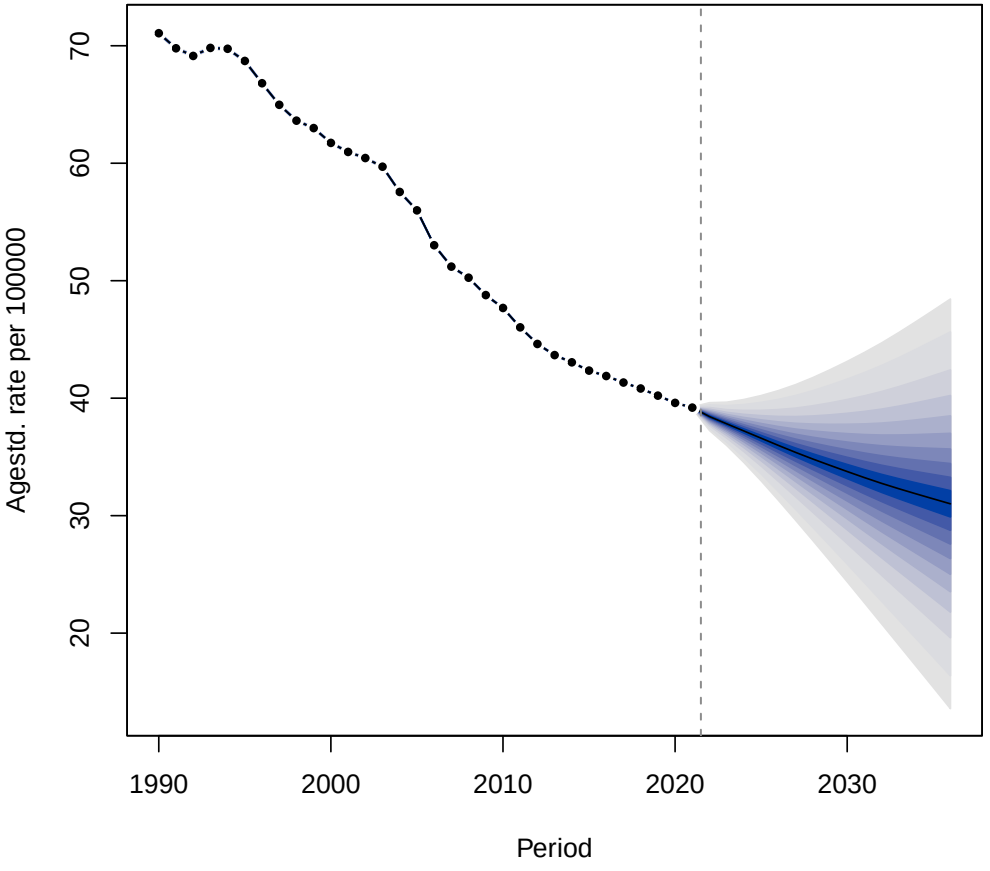
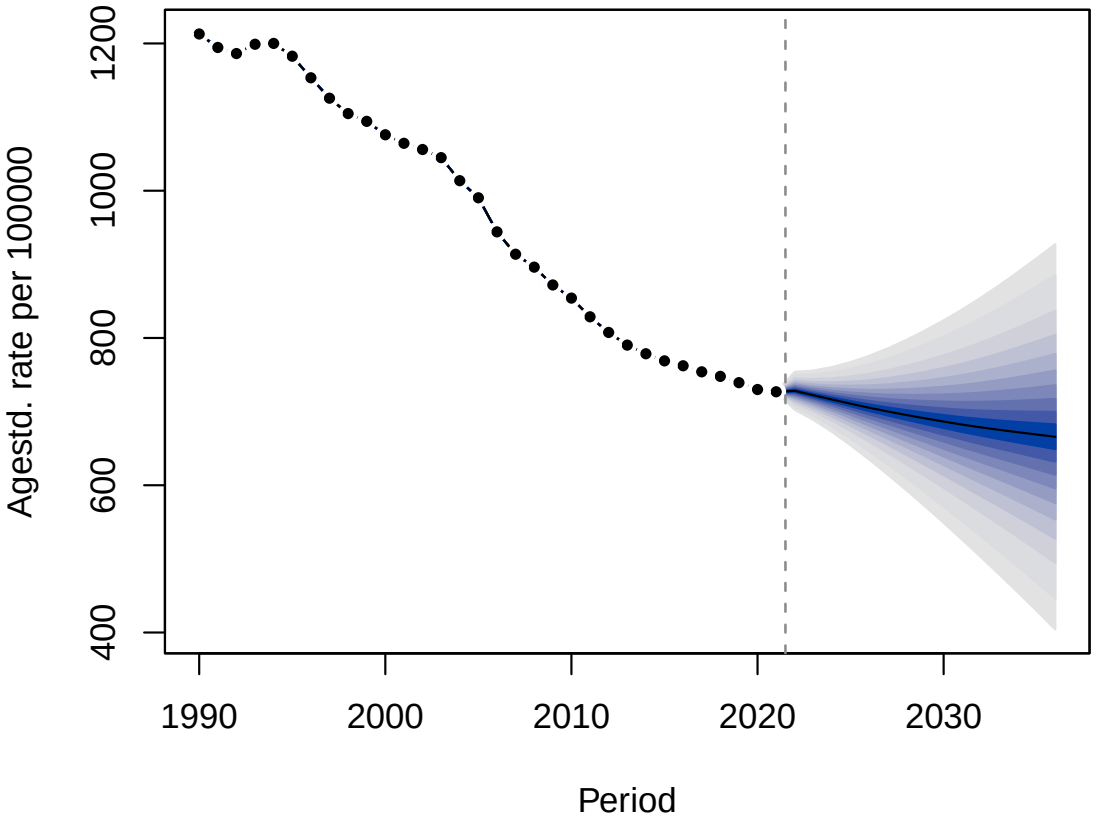

Agestd. rate per 100000
70
60
50
40
30
20
1990
2000
2010
2020
2030
Period

Figure S36: Projected number of new age-standardized DALYs cases for ischemic stroke in females in 2036 according to the SDI.

DALYs = disability-adjusted life years. SDI = sociodemographic index.


Agestd. rate per 100000
400
600
800
1000
1200
1990
2000
2010
2020
2030
Period

Supplementary Methods

1.Query data

Data indicators: GBD database contained all GBD diseases, risk, etiology, injury, etc. Indicators to measure the global burden of disease include death, years of life lost, disability-adjusted life years, prevalence, incidence, and life expectancy. The units of data indicators extracted included number, ratio, percentage, year, probability of death, etc. Annual results for all measures and all GBD age groups from 1990 to 2017 were available for extractable data years. The gender can be selected as male, female, or the combination of the two. The study areas were divided into GBD super regions, regions, countries, selected subnational units, and custom regions (WHO regions, World Bank income levels, etc.).

Search criteria: Click on the "Data tools and practices" drop-down arrow at the top of the GBD database homepage and then select "Data sources". In the new page, click on "VISIT OUR DATA CATALOG". In this interface, Select the data of "GBD 2021 data" or "All IHME data" according to the requirements, click "GBD 2021 data", click "GBD 2019 Data Resources" after seeing the interface, and click "query tool" on the new page. New users can register first and log in directly if they are already registered. The new interface can be divided into two parts, the left side is the search bar, the right side is the result bar, the left side can through GBD Estimate, Measure, Metric, Risk, Cause, Location, Age, Sex, Year of different options to limit the search conditions to obtain their desired data.

2.Data download

Download operation: After obtaining the required data through search, click "Charts" to obtain the trend chart, select "Download" under the left search bar, select "Submit" after the pop-up interface, you can directly select "here" in the new interface, follow the prompts to obtain the desired data, and you can also view your email. Click the link to jump to the download screen.

3.CODEm model calculation method and assumption basis

1. Model calculation method

CODEm (Cause of Death Ensemble model) is an ensemble model used to estimate and predict cause-specific mortality. The calculation method mainly includes the following steps:

Data collection and preprocessing:

Mortality data were collected from different data sources, including national Dynamic registries and oral autopsy reports.

Data were cleaned and preprocessed to remove ambiguous data and sites with alignment gaps.

Covariate selection:

Covariate selection algorithms were used to screen out combinations of covariates associated with a particular cause of death from a large number of possible models.

For example, when studying maternal mortality, 1984 possible models were considered.

Model construction and evaluation:

Multiple models, including mixed-effects linear models and spatio-temporal models, were constructed to account for temporal, geographic, and age correlations.

Each model was evaluated using out-of-sample predictive validity to select the best performing model.

Integrated model construction:

The best performing individual models were combined into an ensemble model (CODEm) to improve the accuracy and stability of prediction.

The performance of the integrated model was evaluated by the root mean square error, the frequency of correctly predicting the time trend, and the 95% coverage of the prediction interval.

2. Model assumption basis

The assumption basis of CODEm model mainly includes the following aspects:

Data distribution assumptions:

Assuming correlations in the data across time, geography, and age groups, spatial-temporal models were used to capture these correlations.

Associations of covariates with cause of death:

Assuming a linear or nonlinear relationship between covariates and cause-specific mortality, a covariate selection algorithm was used to identify the most relevant covariates.

Independence of the model:

It is assumed that the different individual models are independent of each other, and the accuracy of the prediction can be improved by combining these models.

Prediction interval assumptions:

The prediction interval of the model was assumed to cover 95% of the true value, which was used to evaluate the uncertainty of the model.

With these computational methods and assumptions, CODEm models can effectively estimate and predict cause-specific mortality and provide scientific basis for public health decision making.

4. DisMod-MR 2.1 Assumption basis and calculation method

1. Assumptions

DisMod-MR 2.1 is a Bayesian meta-regression tool for estimating epidemiological data on nonfatal diseases. Its assumption basis mainly includes the following aspects:

Data distribution assumptions:

It is assumed that there is an intrinsic link between indicators such as disease incidence, prevalence, remission rate, and case fatality rate, which are not independent.

Bayesian framework:

Bayesian methods were used to deal with the uncertainty and heterogeneity of the data, and the model parameters were estimated through the prior and posterior distributions.

Spatial and temporal correlation:

Assuming correlations between disease data across time, geographic regions, and age groups, spatial-temporal Gaussian process regression models were used to capture these correlations.

Data adjustment:

The meta-regression procedure of Bayesian regularized pruning (MR-BRT) was used to adjust the data bias under different case definitions and research methods, assuming bias between different data sources.

Hierarchy of the model:

It is assumed that the estimates of the model can be progressively refined from the

global level to the national and subnational levels, using a hierarchical model structure to progressively adjust the estimates.

2. Calculation method

The calculation method of DisMod-MR 2.1 mainly includes the following steps:

Data collection and preprocessing:

Epidemiological data were collected from different data sources, including vital registration, oral autopsy reports, population representative surveys and surveillance data.

Data were cleaned and preprocessed to remove ambiguous data and sites with alignment gaps.

Model construction:

The Bayesian meta-regression method was used to construct the model, considering the incidence, prevalence, remission rate and mortality of the disease.

Polynomial network meta-regression, logistic regression, and spatio-temporal Gaussian process regression were used to fit quantities and ratios.

Model fitting and adjustment:

The stochastic Markov chain Monte Carlo (MCMC) method was used to fit the model and estimate the model parameters.

Bayesian regularized pruning (MR-BRT) was used to adjust the bias between different data sources.

Hierarchical estimation:

Starting at the global level and progressively refining to the super-regional, regional, national, and subnational levels, a hierarchical model structure was used to progressively adjust the estimates.

Assessment of uncertainty:

The 95% uncertainty interval (95%UI) of each indicator was estimated to evaluate the uncertainty of the model.

Summary of results:

All subnational level estimates were pooled to the country level, then to the district and super-district levels, and finally to the global level.

With these assumptions and computational methods, DisMod-MR 2.1 can effectively estimate and predict the epidemiological data of non-fatal diseases, and provide scientific basis for public health decision-making.

5. ST-GPR

1.ST-GPR definition

Spatiotemporal Gaussian Process Regression (ST-GPR) is a spatiotemporal Gaussian process regression model that is mainly used to estimate and predict risk factor exposure and mortality in the Global Burden of Disease (GBD) project. It estimates a single indicator by exploiting the correlation between time and space, borrowing the strength of the data between different geographical locations and time points.

2. ST-GPR calculation method

The ST-GPR calculation method is a three-stage process:

Linear mixed-effects model:

A linear mixed-effects model was first run using a set of predictive covariates (e.g., GDP per capita, women's education, and energy consumption). These predicted values provide a general trend of the data.

Spatiotemporal pattern estimates:

In the second step, the spatio-temporal patterns are estimated by applying a series of spatio-temporal weights to average the residuals from the linear model of the first step. These spatio-temporal patterns are added to the linear predictions to generate spatio-temporal predictions.

Gaussian process regression:

Finally, the spatio-temporal prediction is regressed on the data across time as a function of the mean of the Gaussian process regression. The estimated results of the Gaussian process regression serve as the final ST-GPR forecast and generate the full time series estimates for 195 countries from 1990 to 2017.

3. ST-GPR assumption basis

The assumptions underlying ST-GPR include:

Spatial-temporal correlations: Disease data were assumed to be correlated across time, geographic regions, and age groups, and spatial-temporal Gaussian process regression models were used to capture these correlations.

Data adjustment: Assuming bias between different data sources, a meta-regression procedure using Bayesian regularized pruning (MR-BRT) was used to adjust data bias under different case definitions and study methods.

Hierarchical estimation: Assuming that the estimates of the model can be progressively refined from the global level to the national and subnational levels, a hierarchical model structure is used to progressively adjust the estimates.

With these calculation methods and assumptions, ST-GPR can effectively estimate and predict various indicators in the global Burden of Disease project, and provide scientific basis for public health decision-making.

4.BAPC

1. The assumption basis of BAPC

The Bayesian Age-Period-Cohort (BAPC) model is a Bayesian statistical model used to analyze and predict disease burden data, taking into account the effects of Age, Period, and Cohort. Its assumption basis mainly includes the following aspects:

Multiplicative effects of age, period, and cohort:

Suppose that the burden of disease (e.g., incidence, mortality, etc.) can be expressed as a multiplicative effect of age, period, and cohort, namely:

$$\ln(\mathrm{ASR}) = \alpha + \mu_i + \beta_j + \gamma_k$$

Bayesian framework:

Bayesian method was used to estimate the model parameters, and the posterior distribution was obtained by integrating the prior information of the unknown parameters with the sample information.

Random walk model:

A second-order random walk (RW2) model was used to adjust for the effects of age, period, and cohort, assuming that the expected effects of being adjacent in time would be similar.

Integrated nested Laplace approximation (INLA) :

The use of the INLA method to directly approximate the posterior marginal distribution avoids mixing and convergence problems and shows a relatively low error rate.

Prior distribution:

The prior distributions of age, period, and cohort effects were assumed to be inverse gamma distributions, and a second-order random walk model was applied to adjust for overdispersion phenomena.

2. Calculation method of BAPC

The BAPC calculation method is based on the Bayesian statistical framework, and the specific steps are as follows:

Data preparation:

Data on disease burden were collected, including year, age group, incidence, and mortality.

Model construction:

The BAPC package and INLA package in R language were used to build the model. The model can be expressed as follows.

$$\ln(\mathrm{ASR}) = \alpha + \mu_i + \beta_j + \gamma_k$$

Model fitting:

The INLA method was used to fit the model and estimate the effects of age, period, and cohort.

Forecasting:

Use the well-fitted model to make predictions and generate forecasted values of disease burden for future time periods.

Model Evaluation:

The prediction performance of the model is evaluated using indicators such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

Through these assumptions and calculation methods, the BAPC model can effectively analyze and predict disease burden data, providing scientific basis for epidemiological research and public health decision-making.

5.EAPC

EAPC (Estimated Annual Percentage Change) is used to measure the trend of a disease index (such as incidence, mortality, disability-adjusted life years, etc.) in a

specific time period. Its assumption basis mainly includes the following aspects:

Linearity assumption:

Assuming a linear relationship between the natural log value of the disease indicator and time, namely:

$$\ln(\mathrm{ASR}) = \alpha + \beta \times \mathrm{year} + \sigma$$

Where α is the intercept, β is the coefficient of the time variable, and σ is the error term.

Time continuity:

It is assumed that the annual change continues to be stable over the entire period of observation, that is, the trend of change is continuous.

Logarithmic transformation:

Assuming a linear relationship in the change of the disease indicator after natural log transformation helps to stabilize the variance and simplify the model.

2. Calculation method of EAPC

EAPC is calculated based on a linear regression model and the steps are as follows:

Data preparation:

Time series data of disease indicators (such as incidence, mortality, etc.) are collected, usually including year and corresponding index values.

Logarithmic transformation:

The natural logarithmic transformation of disease indicators was calculated as ln(ASR), where ASR is the age-standardized rate.

Linear regression:

Linear regression models were used to fit ln(ASR) as a function of year, namely:

$$\ln(\mathrm{ASR}) = \alpha + \beta \times \mathrm{year}$$

Here, β is the coefficient of the time variable.

Calculate EAPC:

EAPC was calculated from the regression coefficient β using the formula:

$$\mathrm{EAPC} = (e^{\beta} - 1) \times 100$$

Where e is the base of the natural logarithm (approximately 2.718).

Confidence intervals were calculated as follows:

The 95% confidence interval of the EAPC was calculated, usually from the confidence interval of the regression model:

$$\mathrm{CI}_{\mathrm{lower}} = (e^{\mathrm{CI}_{\mathrm{lower}}} - 1) \times 100$$

$$\mathrm{CI}_{\mathrm{upper}} = (e^{\mathrm{CI}_{\mathrm{upper}}} - 1) \times 100$$

Hypothesis testing:

Hypothesis testing on EAPC is usually translated into a t-test for the regression coefficient β, and the test statistic is:

$$t = \frac{\beta}{\mathrm{SE}(\beta)}$$

Where SE(β) is the standard error of β.

We add the ARIMA model for prediction and compare it with the BAPC model, and report the assumption basis and limitations of the two prediction models.

The ARIMA model was used to predict the same outcomes (incidence, prevalence, mortality, and DALYs) from 2021 to 2036, and the results were compared to the projections from the BAPC model. Both models produced consistent trends, suggesting robustness in the long-term predictions.

The Bayesian age-period-cohort (BAPC) model relies on several key assumptions that are critical to the validity and accuracy of the model:

Separability of age, period, and cohort effects

Basic Assumptions: The BAPC model assumes that time-related changes in health outcomes can be separated into age, period, and cohort effects. The age effect refers to the changes in an individual's life course as he or she ages. Period effects reflect the impact of events occurring within a given year on health outcomes; Cohort effect is the variation due to the different characteristics of the population in different birth

years.

The prior distribution of the parameters

Use of prior information: In a Bayesian framework, the BAPC model allows prior information to be incorporated into the parameter estimation process. This means that for an unknown parameter in the model, an appropriate prior distribution needs to be specified for it. For example, a highly concentrated Laplace distribution can be used as a prior for parameters with linear effects, whereas a diffusive normal distribution can be used for slopes of age and period.

Similarity of neighboring effects

Adjacent effect Assumption: The BAPC model assumes that the adjacent effects of age, period, and cohort are similar in time. This means that the model considers the effect changes to be smooth and continuous between adjacent time points or age groups.

Poisson distribution assumption of the data

Poisson distribution: In some cases, the BAPC model assumes that the data follow a Poisson distribution. This assumption applies to count data, such as disease incidence or mortality, etc., and it allows the model to use Poisson regression models when dealing with such data.

These assumptions provide the theoretical basis and computational framework for the BAPC model, but also limit the scope of model application and the interpretation of results. In practice, careful consideration needs to be given to whether these assumptions are applicable to specific research contexts and data characteristics.

The assumptions of the BAPC model have significant impacts on the analysis results. Here is a detailed explanation:

Separability of Age, Period, and Cohort Effects

Identification of Influencing Factors: The assumption that age, period, and cohort effects can be separated allows the model to identify and quantify the roles of these three effects in health outcomes. For example, if the incidence of a certain disease increases over time, the model can isolate the period effect to determine whether this is due to increased diagnoses from medical advancements or an actual rise in disease occurrence due to environmental changes.

Result Interpretation: This assumption enables clearer interpretation of the results by distinctly attributing changes in health outcomes to different factors. For instance, it

can clearly indicate that a high disease incidence in a specific age group is due to physiological changes unique to that age (age effect) or due to particular environmental conditions when that age group was born (cohort effect).

Prior Distributions of Parameters

Robustness of Estimation Results: The choice of prior distributions affects the estimation of parameters. If the prior information is accurate and consistent with the actual data, it can enhance the robustness and accuracy of the estimates. For example, using highly concentrated prior distributions can reduce the uncertainty in parameter estimation, bringing the results closer to the true values.

Subjectivity of Results: The setting of prior distributions may introduce subjectivity, as different priors can lead to different analysis results. This requires researchers to be cautious when setting priors and to provide sufficient justification for the choice of prior distributions.

Similarity of Adjacent Effects

Smoothness Constraint: Assuming similarity between adjacent effects essentially imposes a smoothness constraint on the changes in effects. This causes the model to tend to produce smooth curves in estimating effects, rather than abrupt fluctuations. For example, when estimating age effects, the model assumes that changes in incidence between adjacent age groups are gradual, not sudden jumps.

Reasonableness of Results: In many cases, the smoothness constraint is reasonable because changes in health outcomes are often continuous. However, in some special situations, if the actual effects do indeed change dramatically, this assumption may prevent the model from accurately capturing the true pattern of effects.

Poisson Distribution Assumption for Data

Applicability Limitations: The Poisson distribution assumption is suitable for count data, such as disease incidence or mortality rates. If the data do not conform to the characteristics of the Poisson distribution (such as overdispersion or zero inflation), this assumption can lead to poor model fit, thereby affecting the accuracy of the analysis results.

Bias in Results: When data do not meet the Poisson distribution assumption, it can cause bias in the estimation results. For example, if data exhibit overdispersion, using Poisson regression models may underestimate the variance, thus affecting the significance tests of parameters and the reliability of prediction results.

In summary, the assumptions of the BAPC model significantly impact the analysis

results, helping to identify and interpret the roles of different factors while also potentially limiting the model's applicability and accuracy in certain situations. In practical applications, it is necessary to test and adjust these assumptions based on specific data and research contexts to ensure the reliability and validity of the analysis results.

There are some limitations in the application of Bayesian age-period-cohort (BAPC) prediction model, which are analyzed as follows:

Difficulty in parameter estimation

Identification problems caused by linear relationships: There is a linear relationship between the three factors of age, period and cohort, which makes parameter estimation more difficult. The trend of change estimated using the APC model may even be reversed without certain conditional constraints.

Data dependency

Impact of data quality: The predictive accuracy of the BAPC model depends to some extent on the quality of the input data. If the data is missing, wrong or incomplete, it will affect the fitting effect of the model and the prediction result.

Limitations of data sources: In some studies, data may come from GBD and other databases, but these data themselves are estimated based on system dynamics model and statistical model, which has certain limitations.

Limitations of model assumptions

No support for inclusion of covariates: BAPC model does not support inclusion of covariates in the analysis process, which means that other influencing factors (such as socioeconomic factors and environmental factors) other than age, period and cohort cannot be considered, which limits the ability of the model to explain the complex causes of changes in disease burden.

Computational complexity

High computational resource demand: BAPC model usually adopts Bayesian Markov chain Monte Carlo (MCMC) algorithm for parameter estimation, which has high computational complexity and requires a large amount of computing resources and time.

Limitations of the prediction range

Uncertainty in long-term projections: Although the BAPC model can be used to predict future trends in disease burden, the prediction results of the model may be

more uncertain due to the greater uncertainty of future influencing factors when making long-term projections.

These limitations need to be considered and overcome when applying the BAPC model to improve the accuracy and reliability of the model.

To overcome the problem of data dependence of BAPC models, several strategies can be adopted:

Data quality improvement

Data cleaning and preprocessing: The original data were thoroughly cleaned to remove outliers, duplicate records, and inconsistent data to ensure the accuracy and integrity of the data.

Data calibration: Calibration of data to reduce systematic and random errors by comparison with other data sources or known facts.

Diversity of data sources

Integrating multiple data sources: In addition to traditional medical records and death registration data, multiple data sources such as censuses, socioeconomic surveys, and environmental monitoring can also be integrated to improve data comprehensivity and representativability.

Utilizing alternative data: In cases where some data is missing, alternative data can be sought to fill the gap. For example, sample survey data are used to estimate the overall picture.

Model refinement and validation

Model Hypothesis testing: The hypotheses in the model are rigorously tested to ensure that they are supported in the data. If some assumptions are found not to be true, the model structure should be adjusted in time.

Cross-validation: Cross-validation was used to validate the model and evaluate its performance on different datasets to ensure the stability and generalization ability of the model.

Advantages of the Bayesian approach

Utilization of prior information: The Bayesian approach allows prior information to be incorporated into the model, which can alleviate the problem of insufficient data to some extent. Additional information support for the model is provided by combining expert knowledge and historical data.

Uncertainty quantification: Bayesian methods are able to provide uncertainty quantification of parameter estimates, such as confidence intervals or probability distributions. This helps to assess the impact of data quality on model predictions.

Through these measures, the dependence of BAPC model on data can be effectively reduced, and the accuracy and reliability of model prediction can be improved.

The autoregressive integrated moving average (ARIMA) model is a statistical model widely used in time series analysis to predict future data points. Although the ARIMA model has been successful in many areas, it is based on a series of assumptions and suffers from some limitations. The following are the main assumptions and limitations of the ARIMA model:

The main assumptions of the ARIMA model

Stationarity:

Assumptions: Time series data are stationary, that is, their statistical properties (e.g., mean, variance) remain constant in time.

Limitations: Many real time series data are non-stationary, such as trends and seasonal variations in economic data. If the stationarity assumption is not met by the data, the predictive performance of the model may decrease significantly. Although it is possible to smooth the data through methods such as difference, excessive difference may lead to loss of information.

Linearity:

Assumption: The relationship in time series data is linear, that is, future values can be represented as linear combinations of past values.

Limitations: Many real time series data have nonlinear relationships, such as volatility clustering in financial markets. The inability of ARIMA models to capture these nonlinear features may lead to prediction bias.

Independent identically distributed (IID) error term:

Assumption: The error terms of the model are independently and identically distributed with zero mean and constant variance.

Limitations: In practice, the error term may not satisfy the assumption of independent identically distributed, such as the presence of autocorrelation or heteroscedasticity. This may lead to inaccurate estimation results of the model and affect the reliability of the prediction.

Model Order Selection:

Assumptions: The order of the model (p, d, q) is known or can be determined in some way.

Limitations: Selecting the appropriate model order is a complex process that usually relies on information criteria such as AIC, BIC, or methods such as cross validation. Inappropriate order selection may lead to overfitting or underfitting of the model, affecting the prediction performance.

Limitations of the ARIMA model

Limited capacity for non-stationary data:

The ARIMA model requires the data to be stationary, but many actual data are non-stationary. Although it is possible to smooth the data through methods such as difference, excessive difference may lead to loss of information and affect the predictive ability of the model.

Failure to capture nonlinear relationships:

Based on the assumption of linearity, ARIMA model cannot effectively capture the nonlinear relationship in time series. For data with nonlinear features, ARIMA models may not provide accurate predictions.

Sensitive to outliers:

The ARIMA model is sensitive to outliers in the data, which may lead to inaccurate model parameter estimation and thus affect the prediction results.

Model assumptions are strict:

The assumptions of the ARIMA model are strict, and it is often difficult for the actual data to fully meet these assumptions. For example, the assumption of independent identically distributed error terms may not hold in many cases, leading to inaccurate estimation results of the model.

The forecast is limited:

ARIMA model is mainly used for short-term prediction, but may not be effective for long-term prediction. This is because the model assumes that the statistical properties of the data remain constant over the forecast period, whereas the statistical properties of the actual data may change over time.

Model selection and parameter estimation are complex:

Selecting the appropriate ARIMA model order is a complex process that usually relies

on methods such as information criteria or cross-validation. Inappropriate order selection may lead to overfitting or underfitting of the model, affecting the prediction performance.

Summary

The ARIMA model is a powerful tool for time series analysis, but its assumptions and limitations need to be carefully considered when applied. For data that are non-stationary, nonlinear, or have outliers, it may be necessary to combine other models, such as seasonal ARIMA models, GARCH models, or machine learning methods, to improve the accuracy and reliability of prediction.

Compared with other prediction models, ARIMA model has the following advantages and disadvantages:

Advantages of the ARIMA model

The model is simple, easy to understand and apply: The structure of ARIMA model is relatively simple, and it is mainly modeled by three parts: autoregression (AR), difference (I) and moving average (MA), without the help of other exogenous variables.

High prediction accuracy: For short-term prediction, ARIMA model has high prediction accuracy and reliability, and can better capture the trend and seasonality of time series.

High maturity: ARIMA model has a solid theoretical foundation, and after long-term development and application, it has formed a set of relatively complete theoretical system.

Suitable for non-seasonal data: ARIMA model is suitable for time series data with no obvious seasonal characteristics, and can smooth non-stationary time series by difference operation.

Disadvantages of the ARIMA model

The requirement for data stationarity is high: ARIMA model requires time series to be stationary, or to be stationary after difference operation. If there is an obvious trend or seasonality in the data, it needs to be smoothed first, which may affect the prediction effect of the model.

Parameter selection is complex: ARIMA model needs to determine three parameters p (autoregressive order), d (difference order) and q (moving average order). Improper

parameter selection may lead to overfitting or underfitting of the model.

Only linear relationships can be captured: The ARIMA model can intrinsically only capture linear relationships, and the prediction effect may not be good for data with nonlinear relationships.

The long-term prediction effect is poor: the ARIMA model will gradually decrease the prediction accuracy due to the accumulation of errors in the long-term prediction.

The comparison between ARIMA model and BAPC model in time series prediction is as follows:

ARIMA model

Advantages:

The model is simple, easy to understand and apply: The structure of ARIMA model is relatively simple, and it is mainly modeled by three parts: autoregression (AR), difference (I) and moving average (MA), without the help of other exogenous variables.

High prediction accuracy: For short-term prediction, ARIMA model has high prediction accuracy and reliability, and can better capture the trend of time series.

High maturity: ARIMA model has a solid theoretical foundation, and after long-term development and application, it has formed a set of relatively complete theoretical system.

Suitable for non-seasonal data: ARIMA model is suitable for time series data with no obvious seasonal characteristics, and can smooth non-stationary time series by difference operation.

Disadvantages:

The requirement for data stationarity is high: ARIMA model requires time series to be stationary, or to be stationary after difference operation. If there is an obvious trend or seasonality in the data, it needs to be smoothed first, which may affect the prediction effect of the model.

Parameter selection is complex: ARIMA model needs to determine three parameters p (autoregressive order), d (difference order) and q (moving average order). Improper parameter selection may lead to overfitting or underfitting of the model.

Only linear relationships can be captured: The ARIMA model can intrinsically only capture linear relationships, and the prediction effect may not be good for data with nonlinear relationships.

The long-term prediction effect is poor: the ARIMA model will gradually decrease the prediction accuracy due to the accumulation of errors in the long-term prediction.

The BAPC model

Advantages:

Comprehensive consideration of multiple factors: BAPC model took into account the influence of age, period and cohort on the distribution of NCDS, which could more comprehensively analyze the trends of incidence and mortality of NCDS.

Unique advantages of the Bayesian framework: In the Bayesian framework, BAPC model can use both sample information and prior information to obtain unique parameter estimates, which makes the results more robust and reliable.

Improved prediction accuracy: Compared with the traditional APC model, the BAPC model can obtain more stable and reliable results and improve the prediction accuracy.

Solving the difficulty in parameter estimation: There is a linear relationship among the three factors of age, period, and cohort in the classical APC model, which may lead to difficulty in parameter estimation. The BAPC model effectively solves this problem through the Bayesian approach.

Flexible model selection: The BAPC model allows the flexible selection of different model components and prior probability distributions, which improves the applicability and flexibility of the model.

Advanced computational methods: The BAPC model uses the ensemble nested Laplace approximation (INLA) for posterior marginal distribution approximation, which avoids mixing and convergence problems and shows a relatively low error rate.

Automatic and simplified calculation: The use of the BAPC package for parameter estimation not only does not require complex coding procedures and convergence checks, but also allows flexible selection of age-standardized ratio (ASR) parameters and prior probability distributions to simplify the calculation process.

Predicting future changes: BAPC models can not only describe trends in age, period, and birth cohort, but also predict future changes based on these trends.

Disadvantages:

High computational complexity: The computational process of BAPC model is relatively complex and requires high computational resources and time.

High data requirements: BAPC models require sufficient data to estimate model

parameters, and for cases with a small amount of data, the accuracy and reliability of the model may be affected.

There are many model assumptions: the BAPC model is based on a Bayesian framework and requires assumptions about the prior distribution, which may affect the prediction results of the model.

Comparison

Applicable scenarios:

ARIMA model: Suitable for time series with no or very weak seasonality. The effect was poor when there was obvious seasonality in the data.

BAPC model: suitable for disease prediction that requires comprehensive consideration of age, period and cohort factors, especially in the field of public health and epidemiology.

Predictive power:

ARIMA model: performs well in short-term forecasting, but may not perform well for long-term forecasting and data with nonlinear relationships.

BAPC model: It can more accurately predict the change trend of diseases in different ages, periods and cohorts, and is suitable for long-term prediction and complex data structure.

Model complexity:

ARIMA model: The model structure is relatively simple, with few parameters, which is easy to understand and apply.

BAPC model: The model structure is complex and requires Bayesian estimation and complex calculations.

Data requirements:

ARIMA model: It has high requirements for the stationarity of the data and needs pre-processing such as difference.

BAPC model: Sufficient data are needed to estimate the model parameters, and high data volume and data quality requirements are required.