

**Supplementary Table 1** Baseline characteristics of eligible participants grouped by history of cataract surgery during follow-up period.

Item	Total	Non-cataract at follow-up	Cataract surgery at follow-up	p <sup>a</sup>
Number of participants	207573	184000	23573	-
<b>Demographic characteristics</b>				
Age group				<b>&lt;0.001</b>
45-54	69527 (33.5%)	68038 (37.0%)	1489 (6.3%)	
55-64	70323 (33.9%)	64387 (35.0%)	5936 (25.2%)	
65-74	42334 (20.4%)	32567 (17.7%)	9767 (41.4%)	
≥75	25389 (12.2%)	19008 (10.3%)	6381 (27.1%)	
Gender				0.652
Male	97332 (46.9%)	86311 (46.9%)	11021 (46.8%)	
Female	110241 (53.1%)	97689 (53.1%)	12552 (53.2%)	
Income categories (AUD)				<b>&lt;0.001</b>
<20000	33143 (16.0%)	28523 (15.5%)	4620 (19.6%)	
≥20000-39999	34956 (16.8%)	29807 (16.2%)	5149 (21.8%)	
≥40000-69999	39479 (19.0%)	35409 (19.2%)	4070 (17.3%)	
≥70000	56319 (27.1%)	52545 (28.6%)	3774 (16.0%)	
Missing	43676 (21.0%)	37716 (20.5%)	5960 (25.3%)	
Education level				<b>&lt;0.001</b>
<10 years	21163 (10.2%)	18553 (10.1%)	2610 (11.1%)	
High school/TAFE	130931 (63.1%)	1155492 (62.8%)	15439 (65.5%)	
University or higher	52183 (25.1%)	47069 (25.6%)	5114 (21.7%)	
Missing	3296 (1.6%)	2886 (1.6%)	410 (1.7%)	
Country of birth				0.258
Australia	155537 (74.9%)	137799 (74.9%)	17738 (75.2%)	
Others	50412 (24.3%)	44772 (24.3%)	5640 (23.9%)	
Missing	1624 (0.8%)	1429 (0.8%)	195 (0.8%)	
Residential rurality				<b>&lt;0.001</b>
Major cities	107427 (51.8%)	94602 (51.4%)	12825 (54.4%)	
Inner regional	72629 (35.0%)	64613 (35.1%)	8016 (34.0%)	
Outer regional	21469 (10.3%)	19331 (10.5%)	2138 (9.1%)	
Remote	2041 (1.0%)	1856 (1.0%)	185 (0.8%)	
Missing	4007 (1.9%)	3598 (2.0%)	409 (1.7%)	
Health insurance				<b>&lt;0.001</b>
Private with extras	93980 (45.3%)	83048 (45.1%)	10932 (46.4%)	
Private without extras	24608 (11.9%)	21387 (11.6%)	3221 (13.7%)	
Veterans' Affairs card	3365 (1.6%)	3227 (1.8%)	138 (0.6%)	
Health care concession	47751 (23.0%)	40463 (22.0%)	7288 (30.9%)	
None of the above	33995 (16.4%)	32374 (17.6%)	1621 (6.9%)	

<i>Missing</i>	3874 (1.9%)	3501 (1.9%)	373 (1.6%)	
Marital status				0.157
<i>Married/Partner</i>	159477 (76.8%)	141452 (76.9%)	18025 (76.5%)	
<i>Single/Widowed/Divorced</i>	48096 (23.2%)	42548 (23.1%)	5548 (23.5%)	
Social economic status (SES)				<0.001
<i>1<sup>st</sup> quintile</i>	40717 (19.6%)	36480 (19.8%)	4237 (18.0%)	
<i>2<sup>nd</sup> quintile</i>	41912 (20.2%)	37156 (20.2%)	4756 (20.2%)	
<i>3<sup>rd</sup> quintile</i>	37771 (18.2%)	33385 (18.1%)	4386 (18.6%)	
<i>4<sup>th</sup> quintile</i>	34869 (16.8%)	30966 (16.8%)	3903 (16.6%)	
<i>5<sup>th</sup> quintile</i>	46648 (22.5%)	41022 (22.3%)	5626 (23.9%)	
<i>Missing</i>	5656 (2.7%)	4991 (2.7%)	665 (2.8%)	
Body mass index				<0.001
<i>15-18.4 kg/m<sup>2</sup></i>	2379 (1.1%)	2079 (1.1%)	300 (1.3%)	
<i>18.5-24.9 kg/m<sup>2</sup></i>	73767 (35.5%)	65138 (35.4%)	8629 (36.6%)	
<i>25-29.9 kg/m<sup>2</sup></i>	77461 (37.3%)	68481 (37.2%)	8980 (38.1%)	
<i>≥30 kg/m<sup>2</sup></i>	40574 (19.5%)	36444 (19.8%)	4130 (17.5%)	
<i>Missing</i>	13392 (6.5%)	11858 (6.4%)	1534 (6.5%)	
<b>Chronic diseases and family history</b>				
Diabetes at baseline				<0.001
<i>No</i>	192996 (93.0%)	171691 (93.3%)	21305 (90.4%)	
<i>Yes</i>	14577 (7.0%)	12309 (6.7%)	2268 (9.6%)	
Hypertension at baseline				<0.001
<i>No</i>	141208 (68.0%)	127574 (69.3%)	13634 (57.8%)	
<i>Yes</i>	66365 (32.0%)	56426 (30.7%)	9939 (42.2%)	
Dyslipidemia at baseline				<0.001
<i>No</i>	179966 (86.7%)	160707 (87.3%)	19259 (81.7%)	
<i>Yes</i>	27607 (13.3%)	23293 (12.7%)	4314 (18.3%)	
CVD at baseline				<0.001
<i>No</i>	185249 (89.2%)	165671 (90.0%)	19578 (83.1%)	
<i>Yes</i>	22324 (10.8%)	18329 (10.0%)	3995 (16.9%)	
Cancer at baseline				<0.001
<i>No</i>	178197 (85.8%)	159122 (86.5%)	19075 (80.9%)	
<i>Yes</i>	29376 (14.2%)	24878 (13.5%)	4498 (19.1%)	
Parkinson's disease at baseline				<0.001
<i>No</i>	206695 (99.6%)	183272 (99.6%)	23423 (99.4%)	
<i>Yes</i>	878 (0.4%)	728 (0.4%)	150 (0.6%)	
Anxiety at baseline				<0.001
<i>No</i>	162797 (78.4%)	144606 (78.6%)	18191 (77.2%)	
<i>Yes</i>	15537 (7.5%)	14061 (7.6%)	1476 (6.3%)	
<i>Missing</i>	29239 (14.1%)	25333 (13.8%)	3906 (16.6%)	
Family history of diabetes				<0.001
<i>No</i>	160629 (77.4%)	141922 (77.1%)	18707 (79.4%)	

<i>Yes</i>	46934 (22.6%)	42069 (22.9%)	4865 (20.6%)	
<i>Missing</i>	10 (0.0%)	9 (0.0%)	1 (0.0%)	
Family history of hypertension				<b>&lt;0.001</b>
<i>No</i>	102735 (49.5%)	90060 (48.9%)	12728 (54.0%)	
<i>Yes</i>	104824 (50.5%)	93929 (51.0%)	10842 (46.0%)	
<i>Missing</i>	14 (0.0%)	11 (0.0%)	3 (0.0%)	
Family history of CVD				<b>&lt;0.001</b>
<i>No</i>	90690 (43.7%)	81565 (44.3%)	9125 (38.7%)	
<i>Yes</i>	116872 (56.3%)	102425 (55.7%)	14447 (61.3%)	
<i>Missing</i>	11 (0.0%)	10 (0.0%)	1 (0.0%)	
Family history of stroke				<b>&lt;0.001</b>
<i>No</i>	155383 (74.9%)	138521 (75.3%)	16862 (71.5%)	
<i>Yes</i>	52178 (25.1%)	45468 (24.7%)	6710 (28.5%)	
<i>Missing</i>	12 (0.0%)	11 (0.0%)	1 (0.0%)	
<b>Life style indicators</b>				
Smoking				<b>&lt;0.001</b>
<i>Never smoked</i>	120411 (58.0%)	106364 (57.8%)	14047 (59.6%)	
<i>Ex-smoker</i>	72231 (34.8%)	63574 (34.6%)	8657 (36.7%)	
<i>Current smoker</i>	14850 (7.2%)	13994 (7.6%)	856 (3.6%)	
<i>Missing</i>	81 (0.0%)	68 (0.0%)	13 (0.1%)	
Alcohol drinks				0.611
<i>0</i>	62591 (30.2%)	55384 (30.1%)	7207 (30.6%)	
<i>1-4 drinks/week</i>	42539 (20.5%)	38083 (20.7%)	4456 (18.9%)	
<i>5-7 drinks/week</i>	30641 (14.8%)	26857 (14.6%)	3784 (16.1%)	
<i>7-14 drinks/week</i>	37970 (18.3%)	33659 (18.3%)	4311 (18.3%)	
<i>≥15 drinks/week</i>	30538 (14.7%)	27173 (14.8%)	3365 (14.3%)	
<i>Missing</i>	3294 (1.6%)	2844 (1.5%)	450 (1.9%)	
Physical activities				0.911
<i>0-4 sessions/week</i>	34254 (16.5%)	30613 (16.6%)	3641 (15.4%)	
<i>5-9 sessions/week</i>	60496 (29.1%)	53483 (29.1%)	7013 (29.8%)	
<i>10-14 sessions/week</i>	50494 (24.3%)	44309 (24.1%)	6185 (26.2%)	
<i>≥15 sessions/week</i>	54746 (26.4%)	48917 (26.6%)	5829 (24.7%)	
<i>Missing</i>	7583 (3.7%)	6678 (3.6%)	905 (3.8%)	
Sleeping time				<b>&lt;0.001</b>
<i>&lt;7 hours/day</i>	31005 (14.9%)	27624 (15.0%)	3381 (14.3%)	
<i>≥7-9 hours/day</i>	158514 (76.4%)	140821 (76.5%)	17693 (75.1%)	
<i>≥9 hours/day</i>	12150 (5.9%)	10334 (5.6%)	1816 (7.7%)	
<i>Missing</i>	5904 (2.8%)	5221 (2.8%)	683 (2.9%)	
Sitting time				<b>&lt;0.001</b>
<i>&lt;8 hours/day</i>	145348 (70.0%)	127999 (69.6%)	17349 (73.6%)	
<i>≥8 hours/day</i>	46383 (22.3%)	42304 (23.0%)	4079 (17.3%)	
<i>Missing</i>	15842 (7.6%)	13697 (7.4%)	2145 (9.1%)	
<b>Dietary indicators</b>				

Vegetable intakes				<b>&lt;0.001</b>
<5 serving/day	137343 (66.2%)	122626 (66.6%)	14717 (62.4%)	
≥5 servings/day	64811 (31.2%)	56574 (30.7%)	8237 (34.9%)	
Missing	5419 (2.6%)	4800 (2.6%)	619 (2.6%)	
Fruit intakes				<b>&lt;0.001</b>
<2 serving/day	79881 (38.5%)	71880 (39.1%)	8001 (33.9%)	
≥2 servings/day	116701 (56.2%)	102339 (55.6%)	14362 (60.9%)	
Missing	10991 (5.3%)	9781 (5.3%)	1210 (5.1%)	
Breakfast cereal intakes				<b>&lt;0.001</b>
Non-high fiber	30587 (14.7%)	26864 (14.6%)	3723 (15.8%)	
High fiber	138601 (66.8%)	122576 (66.6%)	16034 (68.0%)	
Missing	38376 (18.5%)	34560 (18.8%)	3816 (16.2%)	
Milk intakes				<b>&lt;0.001</b>
None	10493 (5.1%)	9334 (5.1%)	1159 (4.9%)	
Skimmed fat/reduced fat/soy milk	125262 (60.3%)	110091 (59.8%)	15171 (64.4%)	
Whole milk	67136 (32.3%)	60434 (32.8%)	6702 (28.4%)	
Missing	4682 (2.3%)	4141 (2.3%)	541 (2.3%)	
Red meat intakes				<b>&lt;0.001</b>
0 or 1 serving per week	25567 (12.3%)	23114 (12.6%)	2453 (10.4%)	
2 servings per week	35771 (17.2%)	32135 (17.5%)	3636 (15.4%)	
3 servings per week	18907 (9.1%)	16884 (9.2%)	2023 (8.6%)	
4 servings per week	37460 (18.0%)	33224 (18.1%)	4236 (18.0%)	
≥5 servings per week	54363 (26.2%)	47796 (26.0%)	6567 (27.9%)	
Missing	35505 (17.1%)	30847 (16.8%)	4658 (19.8%)	
Fish intakes				<b>&lt;0.001</b>
1 <sup>st</sup> quartile	14856 (7.2%)	13588 (7.4%)	1268 (5.4%)	
2 <sup>nd</sup> quartile	82800 (39.9%)	73912 (40.2%)	8888 (37.7%)	
3 <sup>rd</sup> quartile	30766 (14.8%)	26800 (14.6%)	3966 (16.8%)	
4 <sup>th</sup> quartile	66957 (32.3%)	58904 (32.0%)	8053 (34.2%)	
Missing	12194 (5.9%)	10796 (5.9%)	1398 (5.9%)	
Cheese intakes				<b>&lt;0.001</b>
1 <sup>st</sup> quintile	14477 (7.0%)	12964 (7.0%)	1513 (6.4%)	
2 <sup>nd</sup> quintile	46570 (22.4%)	41481 (22.5%)	5089 (21.6%)	
3 <sup>rd</sup> quintile	48117 (23.2%)	42876 (23.3%)	5241 (22.2%)	
4 <sup>th</sup> quintile	27141 (13.1%)	24011 (13.0%)	3130 (13.3%)	
5 <sup>th</sup> quintile	57744 (27.8%)	50887 (27.7%)	6857 (29.1%)	
Missing	13524 (6.5%)	11781 (6.4%)	1743 (7.4%)	
Self-rated health				
Self-rated health status				<b>&lt;0.001</b>
Excellent	35900 (17.3%)	32509 (17.7%)	3391 (14.4%)	
Very good	82525 (39.8%)	73005 (39.7%)	9520 (40.4%)	

<i>Good</i>	64499 (31.1%)	56613 (30.8%)	7886 (33.5%)	
<i>Fair</i>	15397 (7.4%)	13683 (7.4%)	1714 (7.3%)	
<i>Poor</i>	1550 (0.7%)	1431 (0.8%)	119 (0.5%)	
<i>Missing</i>	7702 (3.7%)	6759 (3.7%)	943 (4.0%)	
Self-rated quality of life				<b>&lt;0.001</b>
<i>Excellent</i>	35128 (16.9%)	31830 (17.3%)	3298 (14.0%)	
<i>Very good</i>	80775 (38.9%)	71520 (38.9%)	9255 (39.3%)	
<i>Good</i>	62609 (30.2%)	54992 (29.9%)	7617 (32.3%)	
<i>Fair</i>	14621 (7.0%)	12986 (7.1%)	1635 (6.9%)	
<i>Poor</i>	1408 (0.7%)	1303 (0.7%)	105 (0.4%)	
<i>Missing</i>	11577 (5.6%)	10111 (5.5%)	1466 (6.2%)	
<b>Social support/Psychological distress</b>				
Social interaction				<b>&lt;0.001</b>
<i>Low</i>	31907 (15.4%)	29166 (15.9%)	2741 (11.6%)	
<i>Moderate</i>	110543 (53.3%)	99079 (53.8%)	11464 (48.6%)	
<i>High</i>	45758 (22.0%)	38789 (21.1%)	6969 (29.6%)	
<i>Missing</i>	19365 (9.3%)	16966 (9.2%)	2399 (10.2%)	
Depression				<b>&lt;0.001</b>
<i>None</i>	83911 (40.4%)	73156 (39.8%)	10755 (45.6%)	
<i>Mild</i>	67796 (32.7%)	61024 (33.2%)	6772 (28.7%)	
<i>Moderate</i>	26578 (12.8%)	24355 (13.2%)	2223 (9.4%)	
<i>High</i>	10347 (5.0%)	9632 (5.2%)	715 (3.0%)	
<i>Missing</i>	18941 (9.1%)	15833 (8.6%)	3108 (13.2%)	

<sup>a</sup>All P values were calculated using a t-test for continuous variables and the  $\chi^2$  test for categorical variables. Boldface indicates statistical significance, deemed as  $p < 0.05$ .

**Supplementary Table 2** Baseline characteristics of 207,573 participants by residency in the 45-and-up study.

Item	City (Training set)	Regional/rural areas (Validating set)	p <sup>a</sup>
Number of participants	<b>107427</b>	<b>100146</b>	
<b>Demographic characteristics</b>			
Gender			<b>&lt;0.001</b>
<i>Male</i>	51563(48.0%)	45769(45.7%)	
<i>Female</i>	55864(52.0%)	54377(54.3%)	
Ancestry			0.238
<i>Australian</i>	22495(20.9%)	27682(27.6%)	
<i>Europe</i>	25642(23.9%)	19312(19.3%)	
<i>Asia</i>	4332(4.0%)	348(0.3%)	
<i>Mixture</i>	7606(7.1%)	2543(2.5%)	
<i>Mixture</i>	46230(43.0%)	49037(49.0%)	
<i>Missing</i>	1122(1.0%)	1224(1.2%)	
Income categories			<b>&lt;0.001</b>
<20000	15248(14.2%)	17895(17.9%)	
≥20000-39999	15248(14.2%)	19708(19.7%)	
≥40000-69999	19352(18.0%)	20127(20.1%)	
≥70000	34628(32.2%)	21691(21.7%)	
<i>Missing</i>	22951(21.4%)	20725(20.7%)	
Education level			<b>&lt;0.001</b>
<10 years	9500(8.8%)	11663(11.6%)	
<i>High school/TAFE</i>	64878(60.4%)	66053(66.0%)	
<i>University or higher</i>	31353(29.2%)	20830(20.8%)	
<i>Missing</i>	1696(1.6%)	1600(1.6%)	
Country of birth			<b>&lt;0.001</b>
<i>Australia</i>	72434(67.4%)	83103(83.0%)	
<i>Others</i>	34142(31.8%)	16270(16.2%)	
<i>Missing</i>	851(0.8%)	773(0.8%)	
Residential rurality			0.662
<i>Major cities</i>	107427(100%)		
<i>Inner regional</i>		72629(72.5%)	
<i>Outer regional</i>		21469(21.4%)	
<i>Remote</i>		2041(2.0%)	
<i>Missing</i>		4007(4.0%)	
Health insurance			<b>&lt;0.001</b>
<i>Private with extras</i>	55097(51.3%)	38883(38.8%)	
<i>Private without extras</i>	12289(11.4%)	12319(12.3%)	
<i>Veterans' Affairs white or gold card</i>	1914(1.8%)	1451(1.4%)	
<i>Health care concession</i>	21284(19.8%)	26467(26.4%)	

<i>None of the above</i>	14916(13.9%)	19079(19.1%)	
<i>Missing</i>	1927(1.8%)	1947(1.9%)	
Marital status			<b>&lt;0.001</b>
<i>Married/Partner</i>	81046(75.4%)	78431(78.3%)	
<i>Single/Widowed/Divorced</i>	26381(24.6%)	21715(21.7%)	
Social economic status (SES)			<b>&lt;0.001</b>
<i>1<sup>st</sup> quintile</i>	13085(12.2%)	27632(27.6%)	
<i>2<sup>nd</sup> quintile</i>	15982(14.9%)	25930(25.9%)	
<i>3<sup>rd</sup> quintile</i>	17299(16.1%)	20472(20.4%)	
<i>4<sup>th</sup> quintile</i>	19847(18.5%)	15022(15.0%)	
<i>5<sup>th</sup> quintile</i>	40548(37.7%)	6100(6.1%)	
<i>Missing</i>	666(0.6%)	4990(5.0%)	
Body mass index			<b>&lt;0.001</b>
<i>15-18.4 kg/m<sup>2</sup></i>	1286(1.2%)	1093(1.1%)	
<i>18.5-24.9 kg/m<sup>2</sup></i>	39737(37.0%)	34030(34.0%)	
<i>25-29.9 kg/m<sup>2</sup></i>	39793(37.0%)	37668(37.6%)	
<i>≥30 kg/m<sup>2</sup></i>	20080(18.7%)	20494(20.5%)	
<i>Missing</i>	6531(6.1%)	6861(6.9%)	
<b>Chronic diseases and family history</b>			
Diabetes at baseline			0.204
<i>No</i>	99809(92.9%)	93187(93.1%)	
<i>Yes</i>	7618(7.1%)	6959(6.9%)	
Hypertension at baseline			<b>0.001</b>
<i>No</i>	73449(68.4%)	67759(67.7%)	
<i>Yes</i>	33978(31.6%)	32387(32.3%)	
Dyslipidemia at baseline			<b>&lt;0.001</b>
<i>No</i>	92429(86.0%)	87537(87.4%)	
<i>Yes</i>	14998(14.0%)	12609(12.6%)	
CVD at baseline			<b>0.001</b>
<i>No</i>	95642(89.0%)	89607(89.5%)	
<i>Yes</i>	11785(11.0%)	10539(10.5%)	
Cancer at baseline			<b>0.002</b>
<i>No</i>	92466(86.1%)	85731(85.6%)	
<i>Yes</i>	14961(13.9%)	14415(14.4%)	
Parkinson's disease at baseline			0.756
<i>No</i>	107E3(99.6%)	99727(99.6%)	
<i>Yes</i>	459(0.4%)	419(0.4%)	
Anxiety at baseline			<b>&lt;0.001</b>
<i>No</i>	84567(78.7%)	78230(78.1%)	
<i>Yes</i>	8081(7.5%)	7456(7.4%)	
<i>Missing</i>	14779(13.8%)	14460(14.4%)	
Family history of diabetes			0.145

<i>No</i>	82993(77.3%)	77636(77.5%)	
<i>Yes</i>	24428(22.7%)	22506(22.5%)	
<i>Missing</i>	6(0.0%)	4(0.0%)	
Family history of hypertension			<b>&lt;0.001</b>
<i>No</i>	53616(49.9%)	49172(49.1%)	
<i>Yes</i>	53805(50.1%)	50966(50.9%)	
<i>Missing</i>	6(0.0%)	8(0.0%)	
Family history of CVD			<b>&lt;0.001</b>
<i>No</i>	47630(44.3%)	43060(43.0%)	
<i>Yes</i>	59791(55.7%)	57081(57.0%)	
<i>Missing</i>	6(0.0%)	5(0.0%)	
Family history of stroke			0.481
<i>No</i>	80347(74.8%)	75036(74.9%)	
<i>Yes</i>	27074(25.2%)	25104(25.1%)	
<i>Missing</i>	6(0.0%)	6(0.0%)	
<b>Life style indicators</b>			
Smoking			<b>&lt;0.001</b>
<i>Never smoked</i>	63899(59.5%)	56512(56.4%)	
<i>Ex-smoker</i>	36306(33.8%)	35925(35.9%)	
<i>Current smoker</i>	7186(6.7%)	7664(7.7%)	
<i>Missing</i>	36(0.0%)	45(0.0%)	
Number of alcohol drinks			<b>&lt;0.001</b>
<i>0</i>	32785(30.5%)	29806(29.8%)	
<i>1-4 drinks/week</i>	22688(21.1%)	19851(19.8%)	
<i>5-7 drinks/week</i>	15851(14.8%)	14790(14.8%)	
<i>7-14 drinks/week</i>	19489(18.1%)	18481(18.5%)	
<i>≥15 drinks/week</i>	14670(13.7%)	15868(15.8%)	
<i>Missing</i>	1944(1.8%)	1350(1.3%)	
Physical activities			<b>&lt;0.001</b>
<i>0-4 sessions/week</i>	18689(17.4%)	15565(15.5%)	
<i>5-9 sessions/week</i>	32128(29.9%)	28368(28.3%)	
<i>10-14 sessions/week</i>	25904(24.1%)	24590(24.6%)	
<i>≥15 sessions/week</i>	26978(25.1%)	27768(27.7%)	
<i>Missing</i>	3728(3.5%)	3855(3.8%)	
Sleeping time			<b>&lt;0.001</b>
<i>&lt;7 hours/day</i>	17683(16.5%)	13322(13.3%)	
<i>≥7-9 hours/day</i>	81145(75.5%)	77369(77.3%)	
<i>≥9 hours/day</i>	5356(5.0%)	6794(6.8%)	
<i>Missing</i>	3243(3.0%)	2661(2.7%)	
Sitting time			<b>&lt;0.001</b>
<i>&lt;8 hours/day</i>	71092(66.2%)	74256(74.1%)	
<i>≥8 hours/day</i>	28109(26.2%)	18274(18.2%)	
<i>Missing</i>	8226(7.7%)	7616(7.6%)	

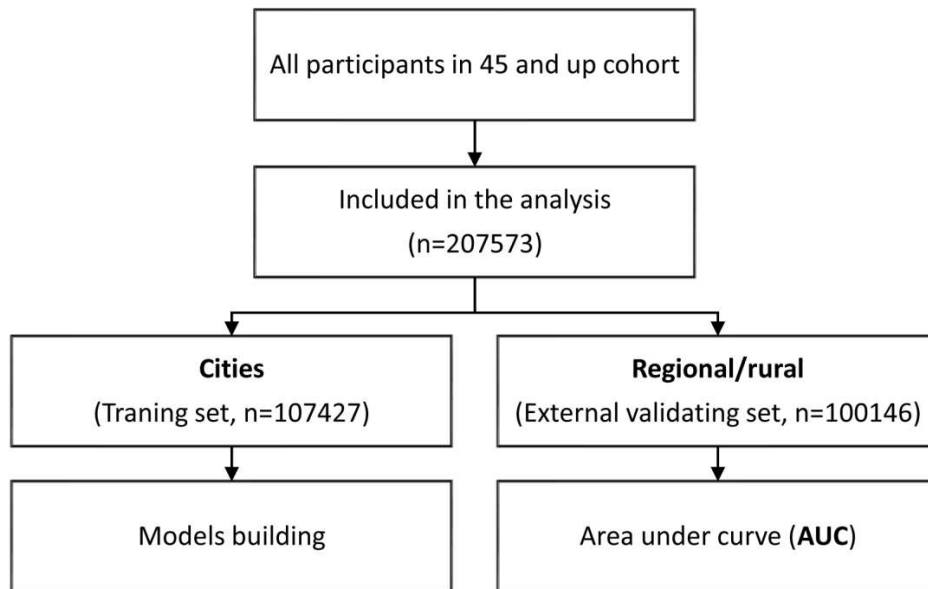


<b>Dietary indicators</b>			
Vegetable intakes			<b>&lt;0.001</b>
<5 serving/day	73446(68.4%)	63897(63.8%)	
≥5 servings/day	30885(28.7%)	33926(33.9%)	
Missing	3096(2.9%)	2323(2.3%)	
Fruit intakes			<b>&lt;0.001</b>
<2 serving/day	40298(37.5%)	39583(39.5%)	
≥2 servings/day	61326(57.1%)	55375(55.3%)	
Missing	5803(5.4%)	5188(5.2%)	
Breakfast cereal intakes			<b>0.003</b>
Non-high fiber	16102(15.0%)	14485(14.5%)	
High fiber	70823(65.9%)	67787(67.7%)	
Missing	20502(19.1%)	17874(17.8%)	
Milk intakes			<b>&lt;0.001</b>
None	5441(5.1%)	5052(5.0%)	
Skimmed fat/reduced fat/soy milk	66817(62.2%)	58445(58.4%)	
Whole milk	32636(30.4%)	34500(34.4%)	
Missing	2533(2.4%)	2149(2.1%)	
Red meat intakes			<b>&lt;0.001</b>
0 or 1 serving per week	5441(5.1%)	5052(5.0%)	
2 servings per week	66817(62.2%)	58445(58.4%)	
3 servings per week	32636(30.4%)	34500(34.4%)	
4 servings per week	2533(2.4%)	2149(2.1%)	
Fish intakes			<b>&lt;0.001</b>
1 <sup>st</sup> quartile	5441(5.1%)	5052(5.0%)	
2 <sup>nd</sup> quartile	66817(62.2%)	58445(58.4%)	
3 <sup>rd</sup> quartile	32636(30.4%)	34500(34.4%)	
4 <sup>th</sup> quartile	2533(2.4%)	2149(2.1%)	
Cheese intakes			<b>&lt;0.001</b>
1 <sup>st</sup> quartile	5441(5.1%)	5052(5.0%)	
2 <sup>nd</sup> quartile	66817(62.2%)	58445(58.4%)	
3 <sup>rd</sup> quartile	32636(30.4%)	34500(34.4%)	
4 <sup>th</sup> quartile	2533(2.4%)	2149(2.1%)	
<b>Self-rated health</b>			.
Self-rated health status			0.188
Excellent	19117(17.8%)	16783(16.8%)	
Very good	41819(38.9%)	40706(40.6%)	
Good	33250(31.0%)	31249(31.2%)	
Fair	8153(7.6%)	7244(7.2%)	
Poor	884(0.8%)	666(0.7%)	
Missing	4204(3.9%)	3498(3.5%)	
Self-rated quality of life			0.545

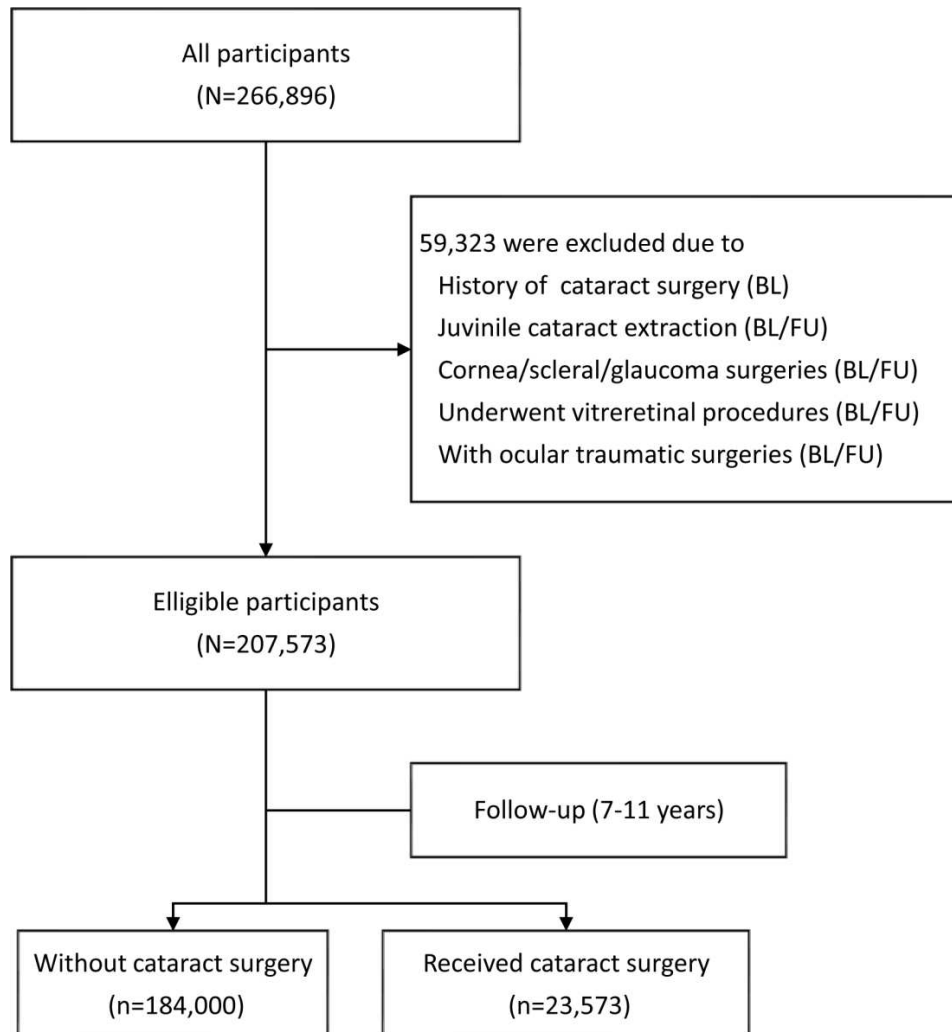
<i>Excellent</i>	18706(17.4%)	16422(16.4%)	
<i>Very good</i>	40984(38.2%)	39791(39.7%)	
<i>Good</i>	32310(30.1%)	30299(30.3%)	
<i>Fair</i>	7750(7.2%)	6871(6.9%)	
<i>Poor</i>	798(0.7%)	610(0.6%)	
<i>Missing</i>	6123(5.7%)	5454(5.4%)	
<b>Social support/Psychological distress</b>			
Depression			<b>0.041</b>
<i>None</i>	43083(40.1%)	40828(40.8%)	
<i>Mild</i>	35348(32.9%)	32448(32.4%)	
<i>Moderate</i>	13928(13.0%)	12650(12.6%)	
<i>High</i>	5399(5.0%)	4948(4.9%)	
<i>Missing</i>	9669(9.0%)	9272(9.3%)	
Social interaction			<b>&lt;0.001</b>
<i>Low</i>	15450(14.4%)	16457(16.4%)	
<i>Moderate</i>	58425(54.4%)	52118(52.0%)	
<i>High</i>	23536(21.9%)	22222(22.2%)	
<i>Missing</i>	10016(9.3%)	9349(9.3%)	

<sup>a</sup>All P values were calculated using a t-test for continuous variables and the  $\chi^2$  test for categorical variables. Boldface indicates statistical significance, deemed as  $p < 0.05$ .

**Supplementary Figure 1** Framework of the external validation (sensitivity analysis by treating city as training set and regional/rural as validation set).



**Supplementary Figure 2** Selection process of eligible participants. BL = baseline; FU = during follow-up



**Supplementary Figure 3** Relative importance of top 10 leading predictors for each predictive model.

