SUPPLEMENTARY MATERIAL

Supplementary methods

To address potential double-counting of individuals surviving to 1 March 2020 in both cohorts, we randomly divided the population into two sub-samples without replacement and cross-validated analysis of RR and IR by selecting the pre-pandemic unexposed group from the first sub-sample and COVID-19 exposed group from the second sub-sample, and vice-versa (**Figure S1.C**), and averaging results (**Table S1, S2**).
We used total contributed time of each patient within 1 year (i.e. time from the start of a period to either death or end of the period) in the survival analysis.

Sensitivity analysis
In sensitivity analyses of KM estimates, stratified by combinations of age, sex, and number of underlying conditions, the best performing model in terms of the estimated versus actual numbers of COVID-19 related deaths was the KM analysis stratified by all three explanatory variables of age, sex, and number of underlying conditions. To internally validate the model against overfitting or underfitting in cross-validated 50% sub-samples (Figure S1.C), we calculated RR and IR on different data fractions (training set) and applied the model on KM results on remaining data (validation set) (**Table S3**).

COVID-19 vaccination started in England in December 2020. To assess vaccination effects on overall RR and IR estimation, we divided the study period into quarters where the 4th quarter (December 2020 to March 2021) included those with 0, 1, or 2 doses of vaccination. We compared RR and IR values of the 4th quarter per vaccination dose to the corresponding quarter in pre-pandemic group (**Table S4**).

Figure S1 Cohort generation for (A) development study; (B) validation study for calculating relative risk (RR) and infection rate (IR); and (C) Cross-validating RR and IR

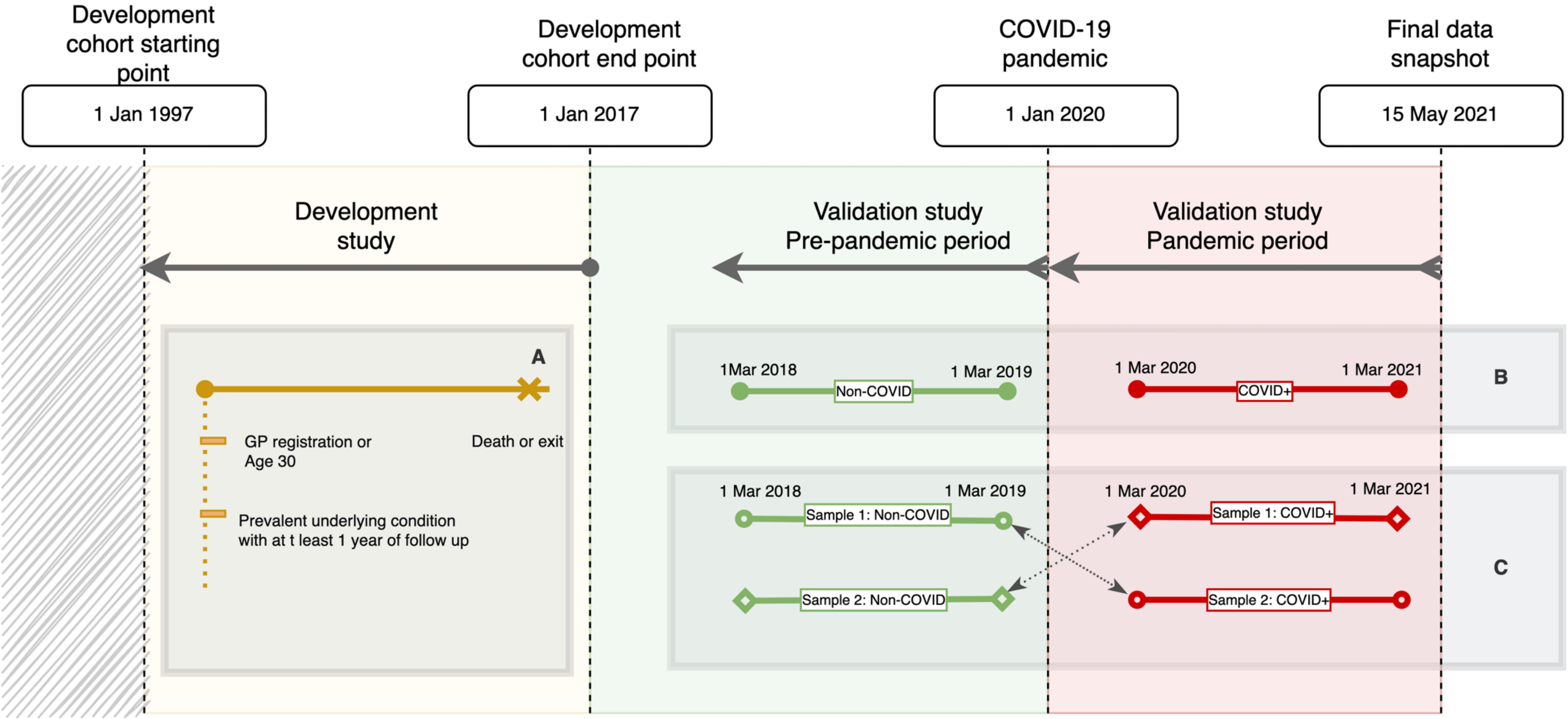
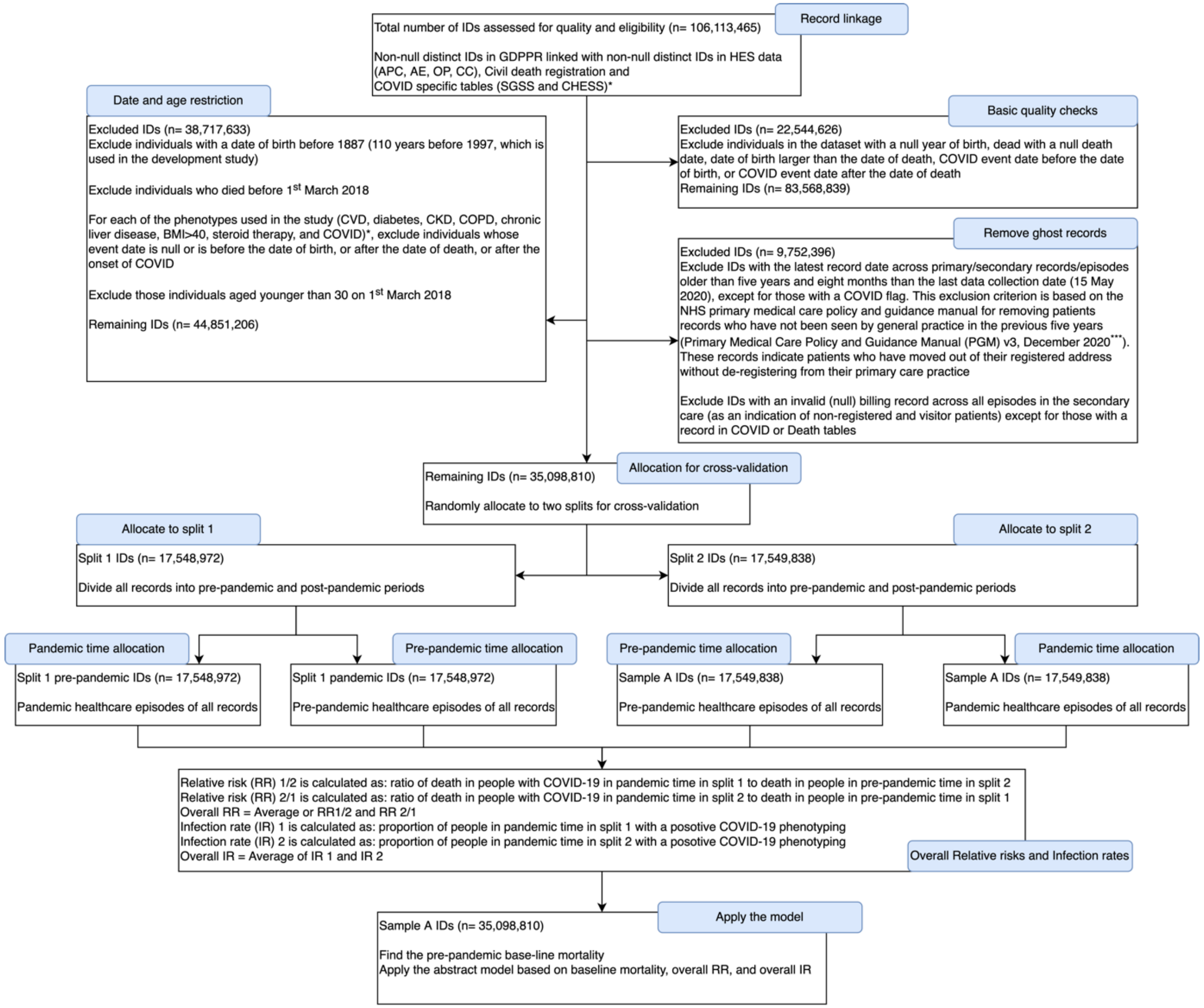


Figure S2 CONSORT diagram of the validation analysis in Trusted Research Environment for England



Total number of IDs assessed for quality and eligibility (n= 106,113,465)

Non-null distinct IDs in GDPPR linked with non-null distinct IDs in HES data (APC, AE, OP, CC), Civil death registration and COVID specific tables (SGSS and CHESS)*

Excluded IDs (n= 38,717,633)
Exclude individuals with a date of birth before 1887 (110 years before 1997, which is used in the development study)

Exclude individuals who died before 1st March 2018

For each of the phenotypes used in the study (CVD, diabetes, CKD, COPD, chronic liver disease, BMI>40, steroid therapy, and COVID)*, exclude individuals whose event date is null or is before the date of birth, or after the date of death, or after the onset of COVID

Exclude those individuals aged younger than 30 on 1st March 2018

Remaining IDs (n= 44,851,206)

Excluded IDs (n= 22,544,626)
Exclude individuals in the dataset with a null year of birth, dead with a null death date, date of birth larger than the date of death, COVID event date before the date of birth, or COVID event date after the date of death
Remaining IDs (n= 83,568,839)

Excluded IDs (n= 9,752,396)
Exclude IDs with the latest record date across primary/secondary records/episodes older than five years and eight months than the last data collection date (15 May 2020), except for those with a COVID flag. This exclusion criterion is based on the NHS primary medical care policy and guidance manual for removing patients records who have not been seen by general practice in the previous five years (Primary Medical Care Policy and Guidance Manual (PGM) v3, December 2020***). These records indicate patients who have moved out of their registered address without de-registering from their primary care practice

Exclude IDs with an invalid (null) billing record across all episodes in the secondary care (as an indication of non-registered and visitor patients) except for those with a record in COVID or Death tables

Remaining IDs (n= 35,098,810)

Randomly allocate to two splits for cross-validation

Split 1 IDs (n= 17,548,972)

Divide all records into pre-pandemic and post-pandemic periods

Split 2 IDs (n= 17,549,838)

Divide all records into pre-pandemic and post-pandemic periods

Split 1 pre-pandemic IDs (n= 17,548,972)

Pandemic healthcare episodes of all records

Split 1 pandemic IDs (n= 17,548,972)

Pre-pandemic healthcare episodes of all records

Sample A IDs (n= 17,549,838)

Pre-pandemic healthcare episodes of all records

Sample A IDs (n= 17,549,838)

Pandemic healthcare episodes of all records

Relative risk (RR) 1/2 is calculated as: ratio of death in people with COVID-19 in pandemic time in split 1 to death in people in pre-pandemic time in split 2
Relative risk (RR) 2/1 is calculated as: ratio of death in people with COVID-19 in pandemic time in split 2 to death in people in pre-pandemic time in split 1
Overall RR = Average or RR1/2 and RR 2/1
Infection rate (IR) 1 is calculated as: proportion of people in pandemic time in split 1 with a posotive COVID-19 phenotyping
Infection rate (IR) 2 is calculated as: proportion of people in pandemic time in split 2 with a posotive COVID-19 phenotyping
Overall IR = Average of IR 1 and IR 2

Sample A IDs (n= 35,098,810)

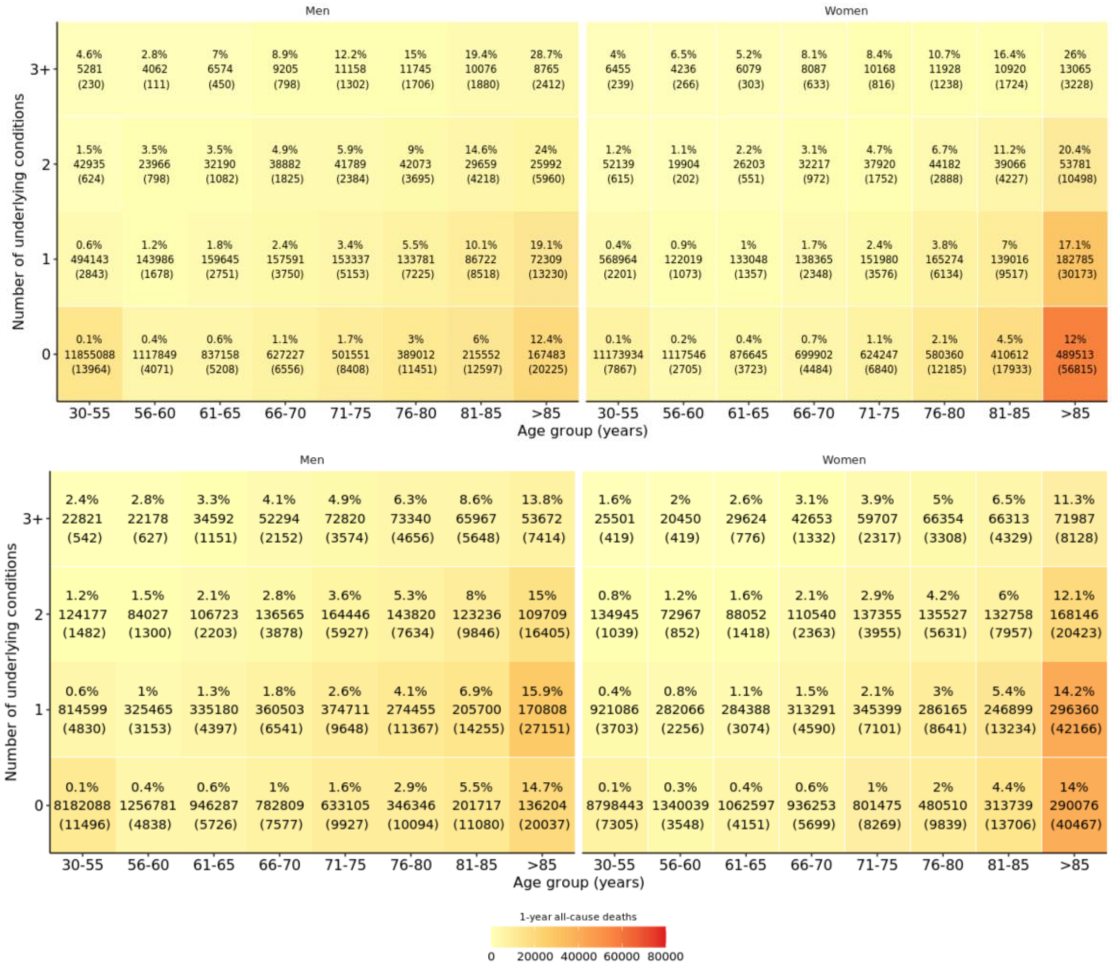
Find the pre-pandemic base-line mortality
Apply the abstract model based on baseline mortality, overall RR, and overall IR

* For more details, refer to the “Data sources” sub-section of methods in the main manuscript·

** For more details, refer to the “Exposures and outcomes of interest” sub-section of methods in the main manuscript·

*** Primary Medical Care Policy and Guidance Manual (PGM) v3.[26]

Figure S3 Baseline one year mortality in England (age ≥ 30) by number of underlying conditions, age category, and sex in development (n=3,862,012 scaled up to mid-2018 population of England of age ≥ 30) and validation cohorts (n=35,098,810)



a) Development cohort (n=3,862,012) scaled up to the mid-2018 population of England aged 30 and over[*]

Number of underlying conditions	Men								Women							
Age group (years)	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85
3+	4.6% 5281 (230)	2.8% 4062 (111)	7% 6574 (450)	8.9% 9205 (798)	12.2% 11158 (1302)	15% 11745 (1706)	19.4% 10076 (1880)	28.7% 8765 (2412)	4% 6455 (239)	6.5% 4236 (266)	5.2% 6079 (303)	8.1% 8087 (633)	8.4% 10168 (816)	10.7% 11928 (1238)	16.4% 10920 (1724)	26% 13065 (3228)
2	1.5% 42935 (624)	3.5% 23966 (798)	3.5% 32190 (1082)	4.9% 38882 (1825)	5.9% 41789 (2384)	9% 42073 (3695)	14.6% 29659 (4218)	24% 25992 (5960)	1.2% 52139 (615)	1.1% 19904 (202)	2.2% 26203 (551)	3.1% 32217 (972)	4.7% 37920 (1752)	6.7% 44182 (2888)	11.2% 39066 (4227)	20.4% 53781 (10498)
1	0.6% 494143 (2843)	1.2% 143986 (1678)	1.8% 159645 (2751)	2.4% 157591 (3750)	3.4% 153337 (5153)	5.5% 133781 (7225)	10.1% 86722 (8518)	19.1% 72309 (13230)	0.4% 568964 (2201)	0.9% 122019 (1073)	1% 133048 (1357)	1.7% 138365 (2348)	2.4% 151980 (3576)	3.8% 165274 (6134)	7% 139016 (9517)	17.1% 182785 (30173)
0	0.1% 11855088 (13964)	0.4% 1117849 (4071)	0.6% 837158 (5208)	1.1% 627227 (6556)	1.7% 501551 (8408)	3% 389012 (11451)	6% 215552 (12597)	12.4% 167483 (20225)	0.1% 11173934 (7867)	0.2% 1117546 (2705)	0.4% 876645 (3723)	0.7% 699902 (4484)	1.1% 624247 (6840)	2.1% 580360 (12185)	4.5% 410612 (17933)	12% 489513 (56815)

b) Validation cohort (n=35,095,810)[*]

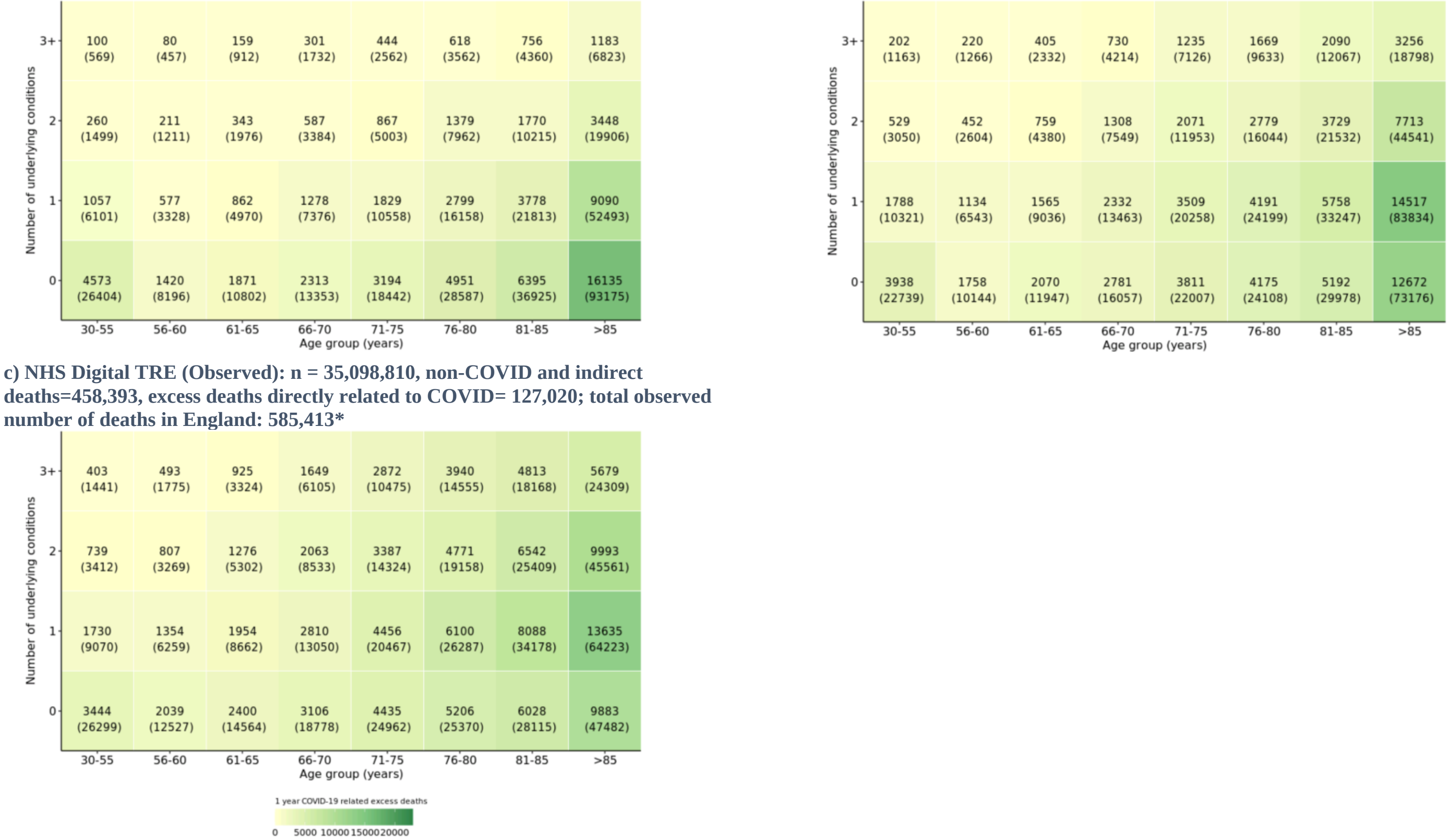
Number of underlying conditions	Men								Women							
Age group (years)	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85
3+	2.4% 22821 (542)	2.8% 22178 (627)	3.3% 34592 (1151)	4.1% 52294 (2152)	4.9% 72820 (3574)	6.3% 73340 (4656)	8.6% 65967 (5648)	13.8% 53672 (7414)	1.6% 25501 (419)	2% 20450 (419)	2.6% 29624 (776)	3.1% 42653 (1332)	3.9% 59707 (2317)	5% 66354 (3308)	6.5% 66313 (4329)	11.3% 71987 (8128)
2	1.2% 124177 (1482)	1.5% 84027 (1300)	2.1% 106723 (2203)	2.8% 136565 (3878)	3.6% 164446 (5927)	5.3% 143820 (7634)	8% 123236 (9846)	15% 109709 (16405)	0.8% 134945 (1039)	1.2% 72967 (852)	1.6% 88052 (1418)	2.1% 110540 (2363)	2.9% 137355 (3955)	4.2% 135527 (5631)	6% 132758 (7957)	12.1% 168146 (20423)
1	0.6% 814599 (4830)	1% 325465 (3153)	1.3% 335180 (4397)	1.8% 360503 (6541)	2.6% 374711 (9648)	4.1% 274455 (11367)	6.9% 205700 (14255)	15.9% 170808 (27151)	0.4% 921086 (3703)	0.8% 282066 (2256)	1.1% 284388 (3074)	1.5% 313291 (4590)	2.1% 345399 (7101)	3% 286165 (8641)	5.4% 246899 (13234)	14.2% 296360 (42166)
0	0.1% 8182088 (11496)	0.4% 1256781 (4838)	0.6% 946287 (5726)	1% 782809 (7577)	1.6% 633105 (9927)	2.9% 346346 (10094)	5.5% 201717 (11080)	14.7% 136204 (20037)	0.1% 8798443 (7305)	0.3% 1340039 (3548)	0.4% 1062597 (4151)	0.6% 936253 (5699)	1% 801475 (8269)	2% 480510 (9839)	4.4% 313739 (13706)	14% 290076 (40467)

[*] Each cell: Mortality risk %, number of people at risk, (number of deaths)

Figure S4 Excess COVID-19 and total deaths over 1 year in England using Lancet 2020 model: (a) CPRD (Predicted), b) NHS Digital TRE (Predicted), and c) NHS Digital TRE (Observed)

a) CPRD (Predicted)†: n=35,407,313 (scaled up from 3,862,012 to mid-2018 population of England aged 30 and over), non-COVID and indirect deaths= 356,186, excess deaths directly related to COVID= 74,628; total predicted number of deaths in England: 430,814*

b) NHS Digital TRE (Predicted)†: n = 35,098,810, non-COVID and indirect deaths=478,971, excess deaths directly related to COVID= 100,338; total predicted number of deaths in England: 579,309*



Number of underlying conditions	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85
3+	100 (569)	80 (457)	159 (912)	301 (1732)	444 (2562)	618 (3562)	756 (4360)	1183 (6823)
2	260 (1499)	211 (1211)	343 (1976)	587 (3384)	867 (5003)	1379 (7962)	1770 (10215)	3448 (19906)
1	1057 (6101)	577 (3328)	862 (4970)	1278 (7376)	1829 (10558)	2799 (16158)	3778 (21813)	9090 (52493)
0	4573 (26404)	1420 (8196)	1871 (10802)	2313 (13353)	3194 (18442)	4951 (28587)	6395 (36925)	16135 (93175)
	Age group (years)							

Number of underlying conditions	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85
3+	202 (1163)	220 (1266)	405 (2332)	730 (4214)	1235 (7126)	1669 (9633)	2090 (12067)	3256 (18798)
2	529 (3050)	452 (2604)	759 (4380)	1308 (7549)	2071 (11953)	2779 (16044)	3729 (21532)	7713 (44541)
1	1788 (10321)	1134 (6543)	1565 (9036)	2332 (13463)	3509 (20258)	4191 (24199)	5758 (33247)	14517 (83834)
0	3938 (22739)	1758 (10144)	2070 (11947)	2781 (16057)	3811 (22007)	4175 (24108)	5192 (29978)	12672 (73176)
	Age group (years)							

c) NHS Digital TRE (Observed): n = 35,098,810, non-COVID and indirect deaths=458,393, excess deaths directly related to COVID= 127,020; total observed number of deaths in England: 585,413*

Number of underlying conditions	30-55	56-60	61-65	66-70	71-75	76-80	81-85	>85
3+	403 (1441)	493 (1775)	925 (3324)	1649 (6105)	2872 (10475)	3940 (14555)	4813 (18168)	5679 (24309)
2	739 (3412)	807 (3269)	1276 (5302)	2063 (8533)	3387 (14324)	4771 (19158)	6542 (25409)	9993 (45561)
1	1730 (9070)	1354 (6259)	1954 (8662)	2810 (13050)	4456 (20467)	6100 (26287)	8088 (34178)	13635 (64223)
0	3444 (26299)	2039 (12527)	2400 (14564)	3106 (18778)	4435 (24962)	5206 (25370)	6028 (28115)	9883 (47482)
	Age group (years)							

†Using observed infection rate over 1 year (IR: 6·27) and observed relative risk on 1-year mortality (RR: 4·34) from NHS Digital TRE

**Each cell: Excess COVID-19 deaths, (Total number of deaths)*

TRE: Trusted Research Environment

Table S1 Cross-validated age-specific and overall infection rate for COVID-19

	Infection rate* (incidence proportion) per 100								
	Age bands for age-specific IR								**Overall**
	30-55	**56-60**	**61-65**	**66-70**	**71-75**	**76-80**	**81-85**	**>85**	-
Whole cohort	7·54 (7·524-7·549)	6·36 (6·337-6·378)	5·16 (5·138-5·188)	3·59 (3·569-3·614)	3·10 (3·080-3·122)	3·69 (3·664-3·718)	5·07 (5·033-5·105)	8·73 (8·681-8·776)	6·27 (6·264-6·280)
Sample 1	7·53 (7·509-7·544)	6·34 (6·307-6·378)	5·18 (5·147-5·218)	3·58 (3·550-3·613)	3·10 (3·070-3·129)	3·70 (3·659-3·734)	5·07 (5·020-5·122)	8·71 (8·643-8·777)	6·27 (6·254-6·277)
Sample 2	7·55 (7·530-7·565)	6·38 (6·346-6·417)	5·14 (5·109-5·179)	3·60 (3·571-3·634)	3·10 (3·074-3·132)	3·69 (3·647-3·723)	5·07 (5·015-5·117)	8·75 (8·680-8·815)	6·28 (6·268-6·291)
Average of two sub-sampled infection rates	7·54 (7·520-7·554)	6·36 (6·327-6·397)	5·16 (5·128-5·199)	3·59 (3·561-3·623)	3·10 (3·072-3·131)	3·69 (3·653-3·728)	5·07 (5·018-5·120)	8·73 (8·662-8·796)	6·27 (6·261-6·284)

*The rate here does not denote the time in the denominator. The denominator is the total number of people at risk at the start of each period.

Table S2 Cross-validated relative risk across two non-overlapping random sub-samples

	Risk ratio (relative risk)								
	Age bands for age-specific								Overall
	30-55	**56-60**	**61-65**	**66-70**	**71-75**	**76-80**	**81-85**	**>85**	-
Non-sampled relative risk (95% CI)	2·35 (2·28-2·42)	3·41 (3·30-3·53)	4·65 (4·52-4·79)	6·79 (6·63-6·95)	7·89 (7·75-8·04)	7·36 (7·25-7·47)	5·65 (5·58-5·72)	2·99 (2·97-3·02)	4·35 (4·32-4·37)
Relative risk of mortality of sample 2 vs sample 1 (95% CI)	2·32 (2·22-2·42)	3·48 (3·31-3·65)	4·63 (4·45-4·82)	6·68 (6·46-6·91)	7·88 (7·68-8·09)	7·44 (7·29-7·60)	5·64 (5·54-5·74)	2·98 (2·94-3·01)	4·34 (4·31-4·38)
Relative risk of mortality of sample 1 vs sample 2 (95% CI)	2·38 (2·28-2·48)	3·35 (3·19-3·52)	4·67 (4·49-4·86)	6·89 (6·67-7·12)	7·90 (7·70-8·11)	7·28 (7·13-7·44)	5·67 (5·57-5·77)	3·01 (2·98-3·047)	4·35 (4·32-4·39)
Average of two sub-sampled relative risks	2·35 (2·250-2·450)	3·41 (3·250-3·585)	4·65 (4·470-4·840)	6·78 (6·565-7·015)	7·89 (7·690-8·100)	7·36 (7·210-7·520)	5·65 (5·555-5·755)	2·99 (2·960-3·029)	4·34 (4·315-4·385)

*All p-values are <0·001

Table S3 Sensitivity analysis and 2-fold cross-validation of relative risk (RR) and population infection rate (IR)

	1-fold validation				**2-fold cross-validation**		**No validation**
					Split1	**Split2**	
Training set percentage	80%	5%	1%	0·01%	50%	50%	100%
Training set size	28,084,615	1,753,218	349,681	3,531	17,065,397	17,549,071	44,913,416
Validation set percentage	20%	95%	99%	99·99%	50%	50%	0
Validation set size	7,014,079	33,345,592	34,749,098	35,095,279	17,549,071	17,065,397	0
Relative risk (Training set)	4·34 (4·31-4·37) p<0·001	4·28 (4·16-4·40) p<0·001	4·43 (4·17-4·70) p<0·001	5·58 (3·12-9·98) p<0·001	4·37 (4·33-4·40) p<0·001	4·33 (4·29-4·37) p<0·001	4·35 (4·32-4·37)
Infection rate (Training set)	6·33 (6·32-6·34)	6·29 (6·24-6·32)	6·28 (6·16-6·32)	6·79 (6·22-7·94)	6·27 (6·26-6·29)	6·34 (6·33-6·35)	6·27 (6·264-6·280)
Estimated excess COVID-19 death (Application of RR and IR from training set on the baseline mortality in validation set)	20,327	93,959	102,199	148,968	50,553	50,603	100,338
Observed COVID-19 death (Validation set)	25,496	120,817	125,719	127,005	63,475	63,747	127,020
Excess/Observed death ratio	0·79	0·77	0·81	1·17	0·79	0·79	0·79
Test aim	Test of large training set	Test of underfitting	Test of underfitting	Test of extreme underfitting	Cross-validation of the 50% splitting	Cross-validation of the 50% splitting	No validation

* **Interpretation:** The training set is used to calculate RR and IR while the validation set is used to estimate baseline 1-year all-cause mortality using KM survival analysis. We applied our model based on RR and IR (from training set) on the baseline mortality (from validation set) to calculate estimated excess COVID-19 deaths. We then compared the estimated excess death to observed excess death in validation set. Due to large number of records in the whole data and randomised splitting without replacement into training and validation sets, even 5% of data as the training set results in close estimations of RR and IR to those of the whole data. To assess any information leak in the analysis pipeline or from training set to validation set, we tested extremely underfitted models. As the size of the training set shrinks below 5%, the training set results in overestimation of RR and IR which demonstrates the lack of information leak from training set to validation set. This also indicates that the similar/close results of RR and IR for training sets beyond 5% is attributable to large data and adequate randomisation. The RR of spilt 1 and split 2 of the 2-fold cross-validation are close to the RR of the whole dataset. In our study, we have used the averages of RR and IRs of cross-validated 50% sub-samples randomised independently of the splits in this table.

Table S4 Assessment of risk ratio, rate ratio, infection risk, and infection rate of different periods based on vaccination in people of age 30 or older

	Non-vaccinated period				Vaccination period				Mixed period
Number of months	1st quarter	2nd quarter	3rd quarter	First 9 months	4th quarter				One year
Calendar months	1 Mar 2020 to 31 May 2020 vs 1 Mar 2018 to 31 May 2018	1 Jun 2020 to 31 Aug 2020 vs 1 Jun 2018 to 31 Aug 2018	1 Sep 2020 to 30 Nov 2020 vs 1 Sep 2018 to 30 Nov 2018	1 Mar 2020 to 30 Nov 2020 vs 1 Mar 2018 to 30 Nov 2018	1 Dec 2020 to 1 Mar 2021 vs 1 Dec 2018 to 1 Mar 2019				1 Mar 2020 to 1 Mar 2021 vs 1 Mar 2018 to 1 Mar 2019
Vaccine dose	-	-	-	-	Overall	0 dose only	1 dose only	2 doses only	Overall
Number of COVID-19*	184567	47264	618895	850726	1290246	843451	96786	2413	2140972
Number of pre-pandemic deaths	122518	108570	114353	345441	133530	133530	133530	133530	478971
Total deaths in pandemic period	175945	110296	127453	413694	171719	134307	35751	1586	585413
Number of deaths with COVID-19*	33163	2219	15669	60016	54802	45186	8198	151	127015
Death-to-exposed (to COVID) ratio	17·968	4·695	2·532	7·055	4·247	5·357	8·470	6·258	5·933
Relative risk (95% CI), p-value	51·47 (50·90-52·06) p<0·0001	15·12 (14·52-15·76) p<0·0001	7·72 (7·59-7·85) p<0·0001	7·17 (7·11-7·23) p<0·0001	11·05 (10·95-11·16) p<0·0001	13·94 (13·80-14·09) p<0·0001	22·05 (21·58-22·52) p<0·0001	16·29 (13·96-19·01) p<0·0001	4·35 (4·32-4·37) p<0·0001
Rate* ratio (95% CI), p-value	55·85 (53·19-58·64) p<0·0001	15·52 (13·83-17·42) p<0·0001	7·76 (7·50-8·03) p<0·0001	7·42 (7·29-7·55) p<0·0001	11·24 (10·98-11·52) p<0·0001	14·28 (13·88-14·69) p<0·0001	22·52 (21·01-24·15) p<0·0001	16·56 (10·58-25·91) p<0·0001	4·42 (4·38-4·46) p<0·0001
Infection risk % (95% CI)	0·54 (0·538-0·543)	0·14 (0·138-0·140)	1·83 (1·824-1·833)	2·49 (2·487-2·498)	3·83 (3·820-3·833)	4·64 (4·626-4·646)	2·94 (2·927-2·944)	1·4 (1·367-1·427)	6·27 (6·264-6·281)

*When the vaccine dose is taken into account, only COVID-19 and deaths after vaccination with the specified dose is included in the analysis·

**The denominator for rate is expressed as person-years·

Interpretation: The low IR values for pre-vaccination period, especially 1Mar-31Aug, which is also evident in https://coronavirus·data·gov·uk/details/cases, could be attributed to data collection and testing methods in the UK. These low numbers affect the overall IR results, causing an underestimation of IR. The values of relative risk or rate ratio in the vaccination period should be interpreted in the context of the study design. In our study, the COVID-unexposed group is from Mar2018-Mar2019 where COVID-19 vaccination was not meaningful; therefore, while the denominator of risk in exposed group decreases (due to narrowing the cohort to specific doses), the denominator of the risk in unexposed group does not change, resulting in large numeric results. One potential approach to address this issue is one-to-one matching between vaccinated exposed people and unexposed people from the pre-pandemic period, which is beyond the scope of our study. However, unlike the quarterly analysis, the overall 1-year RR is based on the total denominator. The ratio of death to the number of exposed to COVID is a better measure to compare different doses of vaccination with regards to COVID-19 infection and mortality. We do not have any information on the actual onset of COVID-19 symptoms, infection date, or immunity level before COVID-19 infection in vaccinated people. The comparison of infection rate between 0, 1 and 2 doses show that people with 1 or 2 doses of vaccine have lower infection rates comparing with people without any vaccine. Although, the inclusion of people with 1 or 2 doses causes a slight decrease in the overall IR (from 4·64 to 3·83) in the 4th quarter, the difference is negligible compared to the effect of low IR in the first stages of the pandemic.

Table S5 Baseline one year mortality risk in England per underlying condition (NHS Digital TRE; n = 35,098,810 age ≥30 years)

Underlying conditions	Age ≤ 70 years			Age > 70 years			All ages		
	N (%)	Observed deaths	1-year mortality risk % (95% CI) *	N (%)	Observed deaths	1-year mortality risk % (95% CI)	N (%)	Observed deaths	1-year mortality risk % (95% CI)
At least one comorbidity except for age > 70	4744687 (13·52)	54497	1·15 (1·14-1·16)	3845654 (10·96)	250715	6·52 (6·49-6·54)	8590341 (24·47)	305121	3·55 (3·54-3·57)
Age > 70	-	-	-	7048826 (20·08)	374134	5·31 (5·29-5·32)	-	-	-
Diabetes	1681745 (4·79)	17363	1·03 (1·02-1·05)	1247763 (3·55)	79630	6·38 (6·34-6·42)	2929508 (8·35)	96993	3·31 (3·29-3·33)
CVD	1382451 (8·93)	23139	1·67 (1·65-1·70)	2049396 (5·84)	171361	8·36 (8·32-8·40)	3431847 (9·78)	194500	5·67 (5·64-5·69)
BMI > 40	934564 (2·66)	24	0·003 (0·002-0·004)	286124 (0·82)	62	0·02 (0·02-0·03)	1220688 (3·48)	86	0·007 (0·006-0·009)
Steroid therapy	1945757 (5·54)	22317	1·15 (1·03-1·16)	1069149 (3·05)	57443	5·37 (5·33-5·42)	3014906 (8·59)	79760	2·65 (2·63-2·66)
COPD	578581 (1·65)	14227	2·46 (2·42-2·50)	701147 (2·00)	64778	9·24 (9·17-9·31)	1279728 (6·65)	79005	6·17 (6·13-6·22)
CKD	502776 (1·43)	4201	0·84 (0·81-0·86)	1204554 (3·43)	40589	3·37 (3·34-3·40)	1707330 (4·86)	44790	2·62 (2·60-2·65)
Chronic liver disease	68596 (0·19)	4405	6·42 (6·24-6·60)	20841 (0·06)	2854	13·69 (13·23-14·16)	89437 (0·25)	57259	8·12 (7·94-8·30)
Number of underlying conditions									
3+	150113 (0·71)	7418	2·97 (2·90-3·03)	530160 (1·51)	39374	7·43 (7·36-7·50)	780273 (2·22)	46792	6·00 (5·94-6·05)
2	857996 (2·44)	14535	1·69 (1·67-1·72)	1114997 (3·18)	77778	6·98 (6·93-7·02)	1972993 (5·62)	92313	4·68 (4·65-4·71)
1	3636578 (10·36)	32544	0·89 (0·89-0·90)	2200497 (6·27)	133563	6·07 (6·04-6·10)	5837075 (16·63)	166107	2·85 (2·83-2·86)
0	23305297 (66·40)	50340	0·22 (0·21-0·22)	3203172 (9·13)	123419	3·85 (3·83-3·87)	26508469 (75·52)	173759	0·655 (0·652-0·659)

* Risk of death per 100 based on Kaplan-Meier estimate of one year mortality
TRE: Trusted Research Environment

Table S6 Overlap between high-risk groups for CVD, DM, CKD, and COPD in England (NHS Digital TRE; n = 35,098,810 age ≥30)

	No CVD	**CVD**	**No DM**	**DM**	**No CKD**	**CKD**	**No COPD**	**COPD**
CVD	-	-	2645479	786368	2802686	629161	2979024	452823
Diabetes	2143140	786368	-	-	2446463	483045	2665508	264000
CKD	1078169	629161	1224285	483045	-	-	1515724	191606
COPD	826905	452823	1015728	264000	1088122	191606	-	-
BMI>40	1030358	190330	909196	311492	1086648	134040	1141735	78953
Chronic liver disease	64493	24944	63228	26209	80570	8867	75949	13488
Steroid therapy	2429993	584913	2607609	407297	2710601	304305	2372332	642574

*CVD: cardiovascular disease; DM: diabetes mellitus; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease