

## Online only supplementary material

Table A 1 List of potential codes for inclusion in a list to identify patients with shortness of breath. 'sob' is the binary variable denoting shortness of breath (1: include; 0: exclude)

medcode	desc	sob	Reason for exclusion
735	[D]Breathlessness	1	
741	[D]Shortness of breath	1	
820	[D]Tachypnoea	0	Abnormally rapid breathing, not breathlessness
982	[D]Apnoea	0	Absence of breathing, not breathlessness
1429	Breathlessness	1	
2506	[D]Sleep apnoea syndrome	0	Absence of breathing, not breathlessness
2563	[D]Respiratory distress	1	
2575	Short of breath on exertion	1	
2737	Respiratory distress syndrome	1	
2931	Difficulty breathing	1	
3007	Newborn transitory tachypnoea	0	Abnormally rapid breathing, not breathlessness
3092	[D]Dyspnoea	1	
4822	Shortness of breath	1	
5175	Breathlessness symptom	1	
5349	Shortness of breath symptom	1	
5896	Dyspnoea - symptom	1	
6326	Breathless - moderate exertion	1	

<b>6434</b>	Paroxysmal nocturnal dyspnoea	1	
<b>7000</b>	O/E - dyspnoea	1	
<b>7534</b>	O/E - respiratory distress	1	
<b>7603</b>	Sleep apnoea	0	Absence of breathing
<b>7683</b>	Breathless - lying flat	1	
<b>7932</b>	Breathless - mild exertion	1	
<b>8148</b>	Obstructive sleep apnoea	0	Absence of breathing
<b>9089</b>	Orthopnoea symptom	1	
<b>9297</b>	[D]Respiratory insufficiency	1	
<b>10114</b>	O/E - tachypnoea	0	Abnormally rapid breathing, not breathlessness
<b>11451</b>	[D]Orthopnoea	1	
<b>12474</b>	SOBOE	1	
<b>18116</b>	Nocturnal dyspnoea	1	
<b>19346</b>	No breathlessness	0	Negation of breathlessness
<b>19426</b>	MRC Breathlessness Scale: grade 3	1	
<b>19427</b>	MRC Breathlessness Scale: grade 2	1	
<b>19429</b>	MRC Breathlessness Scale: grade 5	1	
<b>19430</b>	MRC Breathlessness Scale: grade 4	1	
<b>19432</b>	MRC Breathlessness Scale: grade 1	1	
<b>19576</b>	Apnoea of newborn	0	Absence of breathing
<b>20438</b>	[D]Syndrome sleep apnoea	0	Absence of breathing
<b>20748</b>	Obstructive sleep apnoea	0	Absence of breathing
<b>21801</b>	Breathlessness NOS	1	

<b>22094</b>	Short of breath dressing/undressing	1	
<b>23779</b>	Sleep apnoea	0	Absence of breathing
<b>23924</b>	Scoline apnoea	0	Absence of breathing
<b>24848</b>	Adult respiratory distress syndrome	1	
<b>24889</b>	Breathless - strenuous exertion	1	
<b>26684</b>	Perinatal respiratory distress NOS	0	Related to pregnancy not pathology
<b>26871</b>	Primary sleep apnoea of newborn	0	Absence of breathing
<b>31143</b>	Breathless - at rest	1	
<b>31913</b>	O/E - hyperpnoea	0	Increased depth and rate of breathing, not breathlessness
<b>36301</b>	[D]Hypersomnia with sleep apnoea	0	Absence of breathing
<b>37667</b>	Apnoea alarm monitoring	0	Absence of breathing
<b>37704</b>	O/E - orthopnoea	1	
<b>42287</b>	Borg Breathlessness Score: 6 severe (+)	1	
<b>48539</b>	[D]Insomnia with sleep apnoea	0	Absence of breathing
<b>53771</b>	Dyspnoea on exertion	1	
<b>54594</b>	[M]Mesoblastic nephroma	0	Contains 'sob' string, but is inappropriate
<b>57193</b>	Borg Breathlessness Score: 3 moderate	1	
<b>57283</b>	Introduction of transobturator tape	0	Contains 'sob' string, but is inappropriate

<b>57678</b>	Adult respiratory distress syndrome	1	
<b>57759</b>	Borg Breathlessness Score: 2 slight	1	
<b>57903</b>	CLASP shortness of breath score	1	
<b>58538</b>	Fusobacterial necrotising tonsillitis	0	Contains 'sob' string, but is inappropriate
<b>59860</b>	Borg Breathlessness Score: 4 somewhat severe	1	
<b>60096</b>	CLASP shortness of breath score	1	
<b>64049</b>	Borg Breathlessness Score: 5 severe	1	
<b>65353</b>	Borg Breathlessness Score: 0 none at all	0	Negates breathlessness
<b>67566</b>	Borg Breathlessness Score: 9 very, very sev (almost maximal)	1	
<b>68707</b>	Borg Breathlessness Score: 1 very slight	1	
<b>70061</b>	Borg Breathlessness Score: 7 very severe	1	
<b>70818</b>	Borg Breathlessness Score: 0.5 very, very slight	1	
<b>72334</b>	Borg Breathlessness Score: 8 very severe (+)	1	
<b>72704</b>	[X]Other respiratory distress of newborn	0	Newborn

<b>73978</b>	[X]Other apnoea of newborn	0	Absence of breathing, not breathlessness
<b>93869</b>	Removal of transobturator tape	0	Contains 'sob' string, but is inappropriate
<b>97037</b>	Introduction of transobturator sling	0	Contains 'sob' string, but is inappropriate
<b>98965</b>	Urine beta amino isobutyrate level	0	Contains 'sob' string, but is inappropriate
<b>100177</b>	Berlin questionnaire for sleep apnoea	0	Related to absence of breathing
<b>101843</b>	Borg Breathlessness Score: 10 maximal	1	
<b>Total</b>		78	

Table A 2 Delphi review of codes for shortness of breath. 1: definitely include; 2: uncertain; 3: definitely exclude

medcode	desc	Reviewer 1 Decision	Reviewer 1 comment	Reviewer 2 decision	Reviewer 2 comment
735	[D]Breathlessness	1		1	
741	[D]Shortness of breath	1		1	
1429	Breathlessness	1		1	
2563	[D]Respiratory distress	1		2	I would usually only use this term in children
2575	Short of breath on exertion	1		1	
2737	Respiratory distress syndrome	1		2	I would usually only use this term in children
2931	Difficulty breathing	1		1	
3092	[D]Dyspnoea	1		1	
4822	Shortness of breath	1		1	
5175	Breathlessness symptom	1		1	
5349	Shortness of breath symptom	1		1	
5896	Dyspnoea - symptom	1		1	
6326	Breathless - moderate exertion	1		1	
6434	Paroxysmal nocturnal dyspnoea	1		1	
7000	O/E - dyspnoea	1		1	
7534	O/E - respiratory distress	1		2	
7683	Breathless - lying flat	1		1	
7932	Breathless - mild exertion	1		1	
9089	Orthopnoea symptom	1		1	
9297	[D]Respiratory insufficiency	1		2	

<b>11451</b>	[D]Orthopnoea	1		1	
<b>12474</b>	SOBOE	1		1	
<b>18116</b>	Nocturnal dyspnoea	1		1	
<b>19426</b>	MRC Breathlessness Scale: grade 3	1		1	
<b>19427</b>	MRC Breathlessness Scale: grade 2	1		1	
<b>19429</b>	MRC Breathlessness Scale: grade 5	1		1	
<b>19430</b>	MRC Breathlessness Scale: grade 4	1		1	
<b>19432</b>	MRC Breathlessness Scale: grade 1	1		1	
<b>21801</b>	Breathlessness NOS	1		1	
<b>22094</b>	Short of breath dressing/undressing	1		1	
<b>24848</b>	Adult respiratory distress syndrome	1		2	
<b>24889</b>	Breathless - strenuous exertion	1		1	
<b>31143</b>	Breathless - at rest	1		1	
<b>37704</b>	O/E - orthopnoea	1		1	
<b>42287</b>	Borg Breathlessness Score: 6 severe (+)	1		1	I would never use this scale and not sure who would but if it is coded then it would be relevant
<b>53771</b>	Dyspnoea on exertion	1		1	
<b>57193</b>	Borg Breathlessness Score: 3 moderate	1		1	
<b>57678</b>	Adult respiratory distress syndrome	1		1	
<b>57759</b>	Borg Breathlessness Score: 2 slight	1		1	

<b>57903</b>	CLASP shortness of breath score	1		1	as above
<b>59860</b>	Borg Breathlessness Score: 4 somewhat severe	1		1	
<b>60096</b>	CLASP shortness of breath score	1		1	
<b>64049</b>	Borg Breathlessness Score: 5 severe	1		1	
<b>67566</b>	Borg Breathlessness Score: 9 very, very sev (almost maximal)	1		1	
<b>68707</b>	Borg Breathlessness Score: 1 very slight	1		1	
<b>70061</b>	Borg Breathlessness Score: 7 very severe	1		1	
<b>70818</b>	Borg Breathlessness Score: 0.5 very, very slight	1		1	
<b>72334</b>	Borg Breathlessness Score: 8 very severe (+)	1		1	
<b>101843</b>	Borg Breathlessness Score: 10 maximal	1		1	



Table A 3 Sensitivity analysis: Medical codes used to record shortness of breath in a group of patients in the year before they were diagnosed with lung cancer

medcode	Description	Frequency	Percent	Cum.
<b>4822</b>	Shortness of breath	3,226	24.66	24.66
<b>741</b>	[D]Shortness of breath	1,455	11.12	35.78
<b>1429</b>	Breathlessness	1,116	8.53	44.32
<b>19427</b>	MRC Breathlessness Scale: grade 2	1,106	8.46	52.77
<b>19426</b>	MRC Breathlessness Scale: grade 3	1,010	7.72	60.49
<b>5349</b>	Shortness of breath symptom	816	6.24	66.73
<b>5175</b>	Breathlessness symptom	785	6.00	72.73
<b>19430</b>	MRC Breathlessness Scale: grade 4	764	5.84	78.57
<b>5896</b>	Dyspnoea - symptom	437	3.34	81.91
<b>2575</b>	Short of breath on exertion	415	3.17	85.09
<b>3092</b>	[D]Dyspnoea	395	3.02	88.10
<b>19432</b>	MRC Breathlessness Scale: grade 1	332	2.54	90.64
<b>6326</b>	Breathless - moderate exertion	261	2.00	92.64
<b>19429</b>	MRC Breathlessness Scale: grade 5	189	1.44	94.08
<b>2931</b>	Difficulty breathing	187	1.43	95.51
<b>12474</b>	SOBOE	166	1.27	96.78
<b>7932</b>	Breathless - mild exertion	142	1.09	97.87
<b>735</b>	[D]Breathlessness	66	0.50	98.37
<b>7000</b>	O/E - dyspnoea	49	0.37	98.75
<b>57903</b>	CLASP shortness of breath score	44	0.34	99.08
<b>31143</b>	Breathless - at rest	39	0.30	99.38

<b>7683</b>	Breathless - lying flat	22	0.17	99.55
<b>6434</b>	Paroxysmal nocturnal dyspnoea	19	0.15	99.69
<b>21801</b>	Breathlessness NOS	10	0.08	99.77
<b>11451</b>	[D]Orthopnoea	9	0.07	99.84
<b>9089</b>	Orthopnoea symptom	8	0.06	99.90
<b>24889</b>	Breathless - strenuous exertion	5	0.04	99.94
<b>18116</b>	Nocturnal dyspnoea	3	0.02	99.96
<b>53771</b>	Dyspnoea on exertion	2	0.02	99.98
<b>22094</b>	Borg Breathlessness Score: 10 maximal	1	0.01	99.98
<b>59860</b>	Borg Breathlessness Score: 4 somewhat..	1	0.01	99.99
<b>101843</b>	Short of breath dressing/undressing	1	0.01	100.00
<b>Total</b>	Total	13,081	100.00	

Example R syntax.

Credit: Ben Feakins ([Benjamin.Feakins@phc.ox.ac.uk](mailto:Benjamin.Feakins@phc.ox.ac.uk)).

```
#=====#
#           #
#### R CODE FOR APPENDIX ####
#           #
#=====#

### Set Directory Objects ###
browser.dir <- "/Volumes/PHC/CPRD_data/Browsers"
save.dir <- "~/Desktop"

### Read Data Into R ###
setwd(browser.dir)
medical <- read.delim("medical.txt", header = FALSE,
na.strings = "", stringsAsFactors = FALSE, skip = 1)
names(medical) <- c("medcode", "readcode",
"clinicalevents", "immunisationevents",
"referralevents", "testevents", "readterm",
"databasebuild") # Define headers.

### Define Search Terms ###
search.terms.general <- c("shortness of
breath|sob|pnoea|pnea|puffed|short of
breath|winded|breathless")
search.terms.breath <- c("breath")
search.terms.breath.resid <-
c("short|difficult|labour|labor|distress|insuff")
search.terms.respir <- c("respir")
search.terms.respir.resid <- c("insuff|distress")

### Filtering ###
general <- medical[grepl(search.terms.general,
medical$readterm, ignore.case = TRUE), ]
breath <- medical[grepl(search.terms.breath,
medical$readterm, ignore.case = TRUE), ]
breath <- breath[grepl(search.terms.breath.resid,
breath$readterm, ignore.case = TRUE), ]
respir <- medical[grepl(search.terms.respir,
medical$readterm, ignore.case = TRUE), ]
```

```
respir <- respir[grepl(search.terms.respir.resid,
respir$readterm, ignore.case = TRUE), ]

### Combine Results ###
medcodes <- rbind(general, breath, respir)

### De-Duplicate ###
medcodes <- unique(medcodes)

### Sort ###
medcodes <- medcodes[order(medcodes$medcode), ]

### Remove Codes Not Related to Breathlessness ###
# Further subsetting using grepl().

### Drop Useless Variables ###
medcodes <- medcodes[c("medcode", "readcode")]

### Save a Copy ###
setwd(save.dir)
write.table(medcodes, "sob_library.txt", quote = FALSE,
sep = "\t", na = "", row.names = FALSE)

### Tidying Up ###
rm(browser.dir, save.dir)
rm(search.terms.general, search.terms.breath,
search.terms.breath.resid, search.terms.respir,
search.terms.respir.resid)
rm(medical, general, breath, respir)
rm(medcodes)
```