Supplementary file S1:

Feature importance scores for all 72 features, averaged across the ten best results in hyper-tuning (see Supplementary file S2).

Overall importance 37 weeks	id
0.98758041	S1 DV OECD equivalised income
0.85605563	Partners age at birth of CM
0.84432157	S1 PART Life satisfaction
0.84258112	S1 MAIN Date of Birth (year)
0.83486034	S1 MAIN SOC2000 (without dots)
0.8228174	S1 PART Happy/Unhappy with relationship
0.821527	S1 PART Highest academic qualification
0.82035027	S1 PART Age left full-time education
0.75936859	age left ft education
0.73881838	S1 MAIN Life satisfaction
0.71871111	S1 PART General Health
0.70087589	S1 MAIN Happy/Unhappy with relationship
0.69440132	S1 PART Control over life
0.69067715	S1 DV Number of siblings of CM in household
0.68306262	S1 MAIN Pollution, grime, environmental problems
0.68294946	S1 MAIN Vandalism and damage to property
0.67409882	S1 MAIN Satisfaction with area
0.65324506	S1 PART Longstanding Illness
0.63092511	S1 PART Can run own life
0.61418684	S1 PART Suspects on brink of separation
0.60113396	S1 HHQ Cohort Member Sex C1
0.58678548	preg_bleeding
0.58512601	S1 PART Partner ever used force
0.58143152	S1 PART Depression
0.57925801	MC_IBI_1
0.56637042	S1 MAIN Depression
0.55518964	S1 MAIN Control over life
0.55130603	S1 MAIN Gets what wants out of life
0.54644137	S1 MAIN Suspects on brink of separation
0.52858487	S1 PART Paid job when partner became pregnant
0.52832242	S1 MAIN Any illnesses or problems during pregnancy
0.5165431	eclampsia
0.51454371	S1 MAIN Can run own life
0.50289316	S1 MAIN Type of accommodation
0.50016814	S1 MAIN Damp or condensation
0.49808799	S1 MAIN Respondent's Ethnic Group - 8 category classification
0.49618424	S1 PART Number of cigarettes smoked per day before preg
0.49427377	preg_hyperemesis
0.49091902	Dad not in household
0.45183659	preg_UTI
0.44619761	S1 MAIN Paid job when pregnant

0.43338718	asthma
0.43027782	S1 MAIN Number of cigarettes smoked per day before preg
0.42614891	S1 DV Grandparent of CM in household
0.42031301	S1 PART Often gets in violent rage
0.40715609	S1 MAIN Ever smoked
0.40458956	S1 HHQ Number of cohort babies
0.4034741	S1 MAIN Whether had fertility treatment
0.39221972	preg_non_trivial_infection
0.38841104	S1 MAIN Partner ever used force
0.38469403	depression
0.3733067	S1 MAIN Reading ability - forms
0.3667478	S1 PART Health Conditions: Diabetes
0.35096534	preg_anaemia
0.34717177	thyroid
0.32529832	S1 MAIN Numerical ability - change in shops
0.32252348	hypertension
0.31739328	dermatitis
0.30876654	migraine
0.27980377	diabetes_m
0.2777814	arthritis
0.27613931	epilepsy
0.27516667	endometriosis
0.26443573	Hearing_loss
0.25092891	dorsopathies
0.24985899	anaemia
0.24714923	xxx
0.24388683	irritable_bowel
0.21947669	S1 MAIN Number of units in average day before pregnancy
0.18820304	psoriasis
0.17444372	preg_sciatica
0	S1 PART Frequency of alcohol consumption before preg

Best 10 algorithm average importance scores: 32 weeks

overall	id
0.910889	S1 DV OECD equivalised income
0.853509	Partners age at birth of CM
0.786788	S1 MAIN SOC2000 (without dots)
0.785426	S1 MAIN Date of Birth (year)
0.760617	S1 PART Life satisfaction
0.757794	S1 PART Happy/Unhappy with relationship
0.746222	S1 PART Control over life
0.743651	S1 PART Highest academic qualification
0.736363	age left ft education
0.734127	S1 PART Age left full-time education
0.722538	S1 PART Longstanding Illness
0.711836	S1 PART Can run own life
0.705173	S1 PART Partner ever used force

0.703344	S1 DV Number of siblings of CM in household
0.700166	S1 MAIN Vandalism and damage to property
0.697616	S1 PART General Health
0.696068	S1 MAIN Pollution, grime, environmental problems
0.683107	S1 MAIN Happy/Unhappy with relationship
0.6799	S1 MAIN Life satisfaction
0.675223	S1 PART Depression
0.672835	S1 MAIN Satisfaction with area
0.659699	preg_bleeding
0.63767	S1 PART Often gets in violent rage
0.621317	S1 PART Paid job when partner became pregnant
0.609113	S1 PART Suspects on brink of separation
0.607723	S1 MAIN Control over life
0.593643	S1 MAIN Gets what wants out of life
0.590723	S1 MAIN Respondent's Ethnic Group - 8 category classification
0.590004	S1 HHQ Cohort Member Sex C1
0.588831	S1 MAIN Can run own life
0.578404	S1 MAIN Depression
0.569869	depression
0.558067	S1 MAIN Suspects on brink of separation
0.551809	eclampsia
0.549857	preg_UTI
0.539216	Dad not in household
0.533782	MC_IBI_1
0.531379	S1 MAIN Paid job when pregnant
0.524726	S1 MAIN Damp or condensation
0.517361	S1 DV Grandparent of CM in household
0.511897	S1 MAIN Type of accommodation
0.506997	asthma
0.481643	S1 MAIN Ever smoked
0.480417	preg_non_trivial_infection
0.474636	S1 MAIN Reading ability - forms
0.469636	S1 MAIN Partner ever used force
0.46829	S1 MAIN Number of cigarettes smoked per day before preg
0.467215	S1 PART Number of cigarettes smoked per day before preg
0.455216	Hearing_loss
0.44133	S1 MAIN Whether had fertility treatment
0.439995	S1 MAIN Any illnesses or problems during pregnancy
0.436649	irritable_bowel
0.43472	S1 HHQ Number of cohort babies
0.433575	preg_hyperemesis
0.425048	migraine
0.423814	S1 PART Health Conditions: Diabetes
0.415632	dorsopathies
0.393359	hypertension
0.374417	epilepsy
0.356019	thyroid

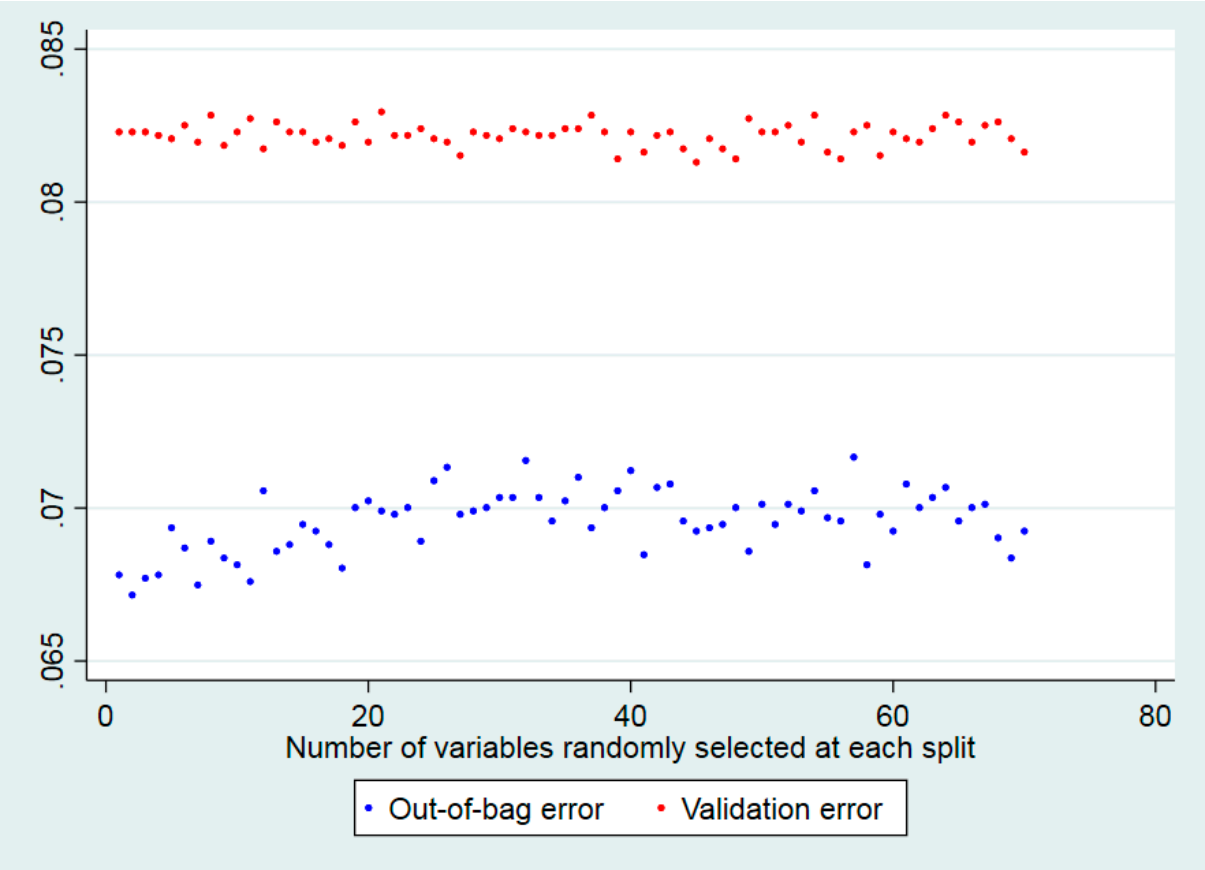
0.355177	S1 MAIN Numerical ability - change in shops
0.313502	diabetes_m
0.2616	preg_anaemia
0.229257	dermatitis
0.211516	S1 MAIN Number of units in average day before pregnancy
0.189147	anaemia
0.164934	arthritis
0.143376	preg_sciatica
0.128683	psoriasis
0.118696	endometriosis
0.087268	xxx
0	S1 PART Frequency of alcohol consumption before preg

Supplementary file S2:

Hyper-tuning results graphs for selecting the number of variables at each split, Stata code including implementation of algorithms with original MCS variable names.

Hyper-tuning results graph for 32 weeks.

Hyper-tuning results graph for 37 weeks.



Stata code (for one set of algorithms): producing importance scores.

```
drop if missing(premie32)

//randomise the data
set seed 1
gen u = uniform()
sort u

rforest premie32 ADOEDE00 part_age ADOTHS00 ADGPAR00 age_left_ed dad_not_pres AMD08E00 MC_IBI_1 Mums_birth dorsopathies preg_sciatica preg_non_trivial_infection preg_anaemia preg_UTI eclampsia preg_hyperemesis preg_bleeding endometriosis arthritis psoriasis dermatitis irritable_bowel asthma hypertension Hearing_loss migraine epilepsy depression xxx diabetes_m thyroid anaemia aplfte00 apacqu00 apwopr00 apwali00 apruli00 apcont00 apforc00 aphare00 apresn00 aprage00 apdrof00 apcipr00 apdean00 apdiab00 apgehe00 aploil00 amarvd00 amarpg00 amarea00 amdamp00 ammoty00 amresn00 amhare00 amforc00 ampuda00 amcipr00 amsmev00 amdean00 amfetr00 ahnoba00 ahcsexa0 ammath00 amform00 amsocc00 amwkpr00 amwali00 amruli00 amcont00 amwant00 amilpr00, type(class) iter(20) numvars(7)

ereturn list

predict pred
```

```
list ahcsexa0 amfetr00 in 1/5

matrix importance = e(importance)

svmat importance

list importance in 1/5
//note that next step will result in all obs being counted as missing
gen id = ""
local mynames : rownames importance
      local k : word count `mynames'
        // If there are more variables than observations
        if `k'>_N {
           set obs `k'
        }
        forvalues i = 1(1)`k' {
           local aword : word `i' of `mynames'
           local alabel : variable label `aword'
           if ("`alabel'"!="") qui replace id= "`alabel'" in `i'
           else qui replace id= "`aword'" in `i'
        }
graph hbar (mean) importance, over(id, sort(1)) bar(1, color(green)) bar(2,
color(red))ytitle(Importance) xsize(6) ysize(11)
```

Example of Stata code for number of cases correctly classified (37 weeks, 9 features)

```
drop if missing(premie37)
set seed 1
gen u = uniform()
sort u
rforest premie37 ADOEDE00 part_age apwali00 Mums_birth amsocc00 aplfte00 apacqu00 aphare00
age_left_ed, type(class) iter(30)
ereturn list
predict p1
list p1 part_age in 1/5
ereturn list
```

Examples of hypertuning code.

```
//hyperparameter tuning for optimal n iterations
drop if missing(premie_258)
set seed 201807
gen u=uniform()
sort u, stable
gen out_of_bag_error1 = .
gen validation_error = .
gen iter1 = .
local j = 0
```

```
forvalues i = 10(5)250 {
local j = `j' + 1
rforest premie_258 ADOEDE00 part_age apwali00 Mums_birth amsocc00 aplfte00 apacqu00
aphare00 age_left_ed amwali00 apgehe00 amhare00 in 1/9100, type(class) iterations(`i') numvars(1)
quietly replace iter1 = `i' in `j'
quietly replace out_of_bag_error1 = `e(OOB_Error)' in `j'
predict p in 9100/18201
replace validation_error = `e(error_rate)' in `j'
drop p
}
label variable out_of_bag_error1 "Out-of-bag error"
label variable iter1 "Iterations"
label variable validation_error "Validation error"
scatter out_of_bag_error1 iter1, mcolor(blue) msize(tiny) ||
scatter validation_error iter1, mcolor(red) msize(tiny)

///hyperparameter tuning for n vars at each split
generate oob_error = .
generate nvars = .
generate val_error = .
local j = 0
forvalues i = 1(1)9 {
local j = `j' + 1
rforest premie_258 ADOEDE00 part_age apwali00 Mums_birth amsocc00 aplfte00 apacqu00
aphare00 age_left_ed in 1/9100, type(class) iter(30) numvars(`i')
quietly replace nvars = `i' in `j'
quietly replace oob_error = `e(OOB_Error)' in `j'
predict p in 9100/18201
quietly replace val_error = `e(error_rate)' in `j'
drop p
}
label variable oob_error "Out-of-bag error"
label variable val_error "Validation error"
label variable nvars "Number of variables randomly selected at each split"
scatter oob_error nvars, mcolor(blue) msize(tiny) || scatter val_error nvars, mcolor(red) msize(tiny)

frame put val_error nvars, into(development)
frame development {
sort val_error, stable
local min_val_err = val_error[1]
local min_nvars = nvars[1]
}
display "Minimum Error: `min_val_err´; Corresponding number of variables `min_nvars´"
frame drop development
```