Derivation of a prediction model for a diagnosis of depression in young adults: a matched case-control study using electronic primary care records.

Nichols, Linda; Ryan, Ronan; Connor, Charlotte; Birchwood, Maximillian; Marshall, Tom

DOI:
10.1111/eip.12332

Title

Derivation of a prediction model for depression in young adults: a matched case-control study using electronic primary care records

Authorship

1 Tom Marshall, Professor of Public Health and Primary Care. [1]

E-mail: T.P.Marshall@bham.ac.uk

Tel: 0121 414 7422

2 Linda Nichols, Research Fellow in Statistics. [1]

E-mail: L.Nichols@bham.ac.uk

3 Ronan Ryan, Research Fellow. [1]

E-mail: R.P.Ryan@bham.ac.uk

4 Charlotte Connor, Senior Research Fellow,. [3]

E-mail: Charlotte.Connor@bsmhft.nhs.uk

5 Max Birchwood, Professor of Clinical Psychology. [2]

E-mail: M.J.Birchwood@warwick.ac.uk

1 School of Health and Population Sciences, University of Birmingham, Edgbaston, Birmingham, UK, B15 2TT

2 Mental Health & Wellbeing, University of Warwick, Coventry, UK, CV4 7AL

3 Birmingham & Solihull Mental Health Foundation Trust, Birmingham, UK.

Contribution:

TM developed the original idea in discussion with CC and MB. Data extraction and initial analysis was led by RR and LN with final data analyses undertaken by LN. All authors contributed to writing the paper.

Correspondence:

Correspondence should be addressed to.

Statistics:

Word count (abstract):	150
Word count (main text):	3271
Tables:	3
Figures	2
References:	49

Abstract

Background

Approximately 80,000 children and young people in the UK suffer from severe depression but many are untreated due to poor identification of early warning signs and risk factors.

Aims

Derive and investigate discrimination characteristics of a prediction model for a first diagnosis of depression in young people aged 15-24 years.

Method

A matched case control study, using electronic primary care records. Stepwise conditional logistic regression modelling investigated 42 potential predictors including symptoms, co-morbidities, social factors, drug and alcohol misuse.

Results

Of the socioeconomic and symptomatic predictors identified, the strongest associations were with depression symptoms and other psychological conditions. School problems and social services involvement were prominent predictors in males aged 15 to 18 years, work stress in females aged 19 to 24 years.

Conclusion

Our model is a first step in the development of a predictive model identifying early warning signs of depression in young people in primary care.

Ethical approval

Research using THIN is approved by the NHS South-East Multi-centre Research Ethics Committee (MREC) in 2003, subject to review by an independent scientific review committee. This project was approved by Scientific Review Committee on 3rd Oct 2014. (SRC Ref: 14-056)

Declaration of Interest

None

Funding

TM, MB and CC are partially funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care West Midlands. This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health.

Background

Approximately 80,000 children and young people in the UK are believed to suffer from severe depression including 8,000 aged under 10 years. [1] A meta-analysis of data from 60,000 adolescents, suggests that the point prevalence for major depression disorder for young people aged between 13 to 18 years is around 6%, [2] international research suggest that between 3-9% of adolescents meet the criteria for depression at any given point during their adolescence, with a lifetime prevalence of up to 20%. [3, 4, 5, 6] Poor outcomes exist for such young people and increased likelihood of behavioural problems, poor functioning, greater chance of substance misuse, and attempted or completed suicide.[7] Those experiencing one episode of depression are also at increased risk of recurrence and of their depression continuing into adult life. [8, 9] Duration of episode (more than 6-months) has also been found to determine the likelihood of recurrent episodes of depression and anxiety during adulthood making a good case for early intervention as a prevention strategy for longstanding mental health problems.[10]

The Royal College of Paediatrics and Child Health emphasises the importance of early intervention for both mental and physical health.[11] This will have far reaching consequences for the future well-being of young people and important economic implications: early intervention and prevention strategies argued to be excellent value for money with a broad range of additional benefits.[12] However, presently in the UK, a 'late intervention' approach persists, which is costly and has little impact on the emotional well-being of young people. [13] In light of this, greater understanding of risk factors associated with the development of depression in young people has become a healthcare priority. The UK Department of Health independent review of Child & Adolescent Mental Health Services (CAMHS), has emphasised the necessity for universal services, such as Primary Care and School Nursing, to improve their understanding of the likely precursors to depression and emotional disorders with the aim of improving outcomes for young people. [14]

A survey of 11,154 young people in Norway, found only a third of those aged 15–16 years, reported seeking early professional help for their anxiety and depression. [15] Reluctance to seek help is often due to fears of stigmatisation or concerns about confidentiality. [16] However, even when help *is* sought by a young person, limited appointment times and a propensity for consultations to focus on physical symptoms can result in mental health issues being missed or going unrecognised. Rates of recognition by healthcare professionals are as low as 18% in some US studies. [17] Raising awareness of that depression should be considered as a diagnosis may help.

Screening tools for depression do not offer a solution. A review of the effectiveness of screening for child and adolescent depression in primary care settings[18] concluded that the evidence base for present-day screening tools, such as the Patient Health Questionnaire for Adolescents (PHQ-A), the Beck Depression Inventory-Primary Care Version (BDI-PC) and the Strengths & Difficulties questionnaire (SDQ), was limited. [19, 20, 21] Great variations in sensitivity are reported with these tools, few are tested with large sample sets or with younger children they are generally only used when depression was already suspected because of the presence of indicators such as antisocial behaviour, diminished school performance, social withdrawal, substance abuse or behavioural difficulties[22]. These indicators are useful but routine consideration of additional factors may also be helpful. An evidence review in 2010, [23] revealed a wide range of factors associated with the development of depression, including somatic symptoms, such as physical health [24] and sleeping problems [25] with an incremental association observed between *number* of somatic complaints and *severity* of depression in young people (16-17 years of age).[26] Smoking behaviour, [27, 28] often related to socio-

economic status has also been argued to precede the onset of depression,[29, 30, 31] rather than simply being a function of it. Such findings highlight the complex nature of depression and how recognition of it may be masked by a variety of factors within a primary care setting.

Prediction models have also been developed for anxiety and for depression in adults in primary care.[32, 33] These have good discrimination characteristics but their practical utility is limited by requiring information not normally available to general practitioners (Short Form 12 scores). Electronic primary care records include a vast amount of electronic information on symptomatology and other patient characteristics which may assist in identifying young people at risk of developing depression. Successful prediction models using such records have been derived to identify patients likely to develop conditions such as cancer[34] or diabetes[35] and those likely to be admitted to hospital as emergencies. [36] It is not known, however, whether an equivalent model could also be used to predict a diagnosis of depression young people.

This study aims to derive and investigate the discrimination characteristics of a prediction model for a diagnosis of depression in young adults aged between 15 to 24 years. The objective is to determine which recorded symptoms, diagnoses and additional individual characteristics may contribute to a future prediction model. If successful this may lead to further development of a prediction model for diagnosis of depression.

Methods

Study design & Setting

A matched case control study was undertaken using The Health Information Network database (THIN): a large dataset of anonymised electronic medical records extracted from general practices using Vision medical records software.[37] In March 2014 THIN included data from 3.7 million patients currently enrolled with 578 general practices across the UK. The population is broadly representative of the UK population although it includes slightly fewer persons aged under 25 years than the general population.[38] Data include administrative details such as date of entry and departure from the database; demographic details and postcode related deprivation index (Townsend quintile); symptoms, diagnoses, prescriptions and laboratory test results. Research using THIN is approved by the NHS South-East Multi-centre Research Ethics Committee (MREC) in 2003 subject to review by an independent scientific review committee.[39]

Practices were included if they had contributed at least one year of data after the latest of three dates: practice acceptable mortality reporting date,[40] the start of the study period and the date the practice started using Vision software. The study period was defined as between 1st January 2000 and 21st December 2012.

Participants

Cases were young people aged between 15 to 24 (from mid to late adolescence) with an incident first diagnosis of depression within at least six months of registration with the practice (i.e. six months' observation) prior to diagnosis. This age range was chosen because fifteen years is considered to be mid-adolescence; recent research revealing neurological changes in the brain continue through to mid-twenties.[41]

Incident depression was defined as the first occurrence of any of a list of clinical codes (Read codes[42]) for depression or a first prescription for an anti-depressant drug from the appropriate section of the British National Formulary.[43] Drugs included tricyclic and related antidepressant drugs, monoamine oxidase inhibitors, selective serotonin reuptake inhibitors or other antidepressant drugs. Date of diagnosis was the index date.

Exclusion criteria included patients in whom the first clinical code for depression indicated a history of depression, (implying a previous diagnosis) and patients a first diagnosis of depression aged younger than 15 years.

Each case was matched on practice, index date, gender and age (up to ± 3 years) to three controls, selected without replacement. Eligible controls had no diagnosis of depression up until index date of their matched case. This means that a case could also be a control if they had not experienced depression up until the index date of their matched case. Controls could also have a diagnosis of depression after age 24.

Exposures / Variables

Exposure variables include: Townsend deprivation quintile; symptoms of depression; somatic symptoms linked to depression; co-morbidities (chronic diseases); family and social factors; drug and alcohol misuse; other psychological conditions.

Depression symptoms include anxiety, low mood, tiredness, loss of enjoyment, too little sleep, too much sleep, eating disorders, weight loss, weight gain, bed wetting, excessive sweating or self-harm.

Somatic symptoms include headache, dyspepsia, dysmenorrhea, abdominal pain, back pain, ill-defined conditions, frequency of consultation (for any reason) and other somatic symptoms.

Co-morbidities include diabetes, epilepsy, asthma, early or late puberty and skin problems.

Family and social factors include: childhood emotional problems, divorce, homelessness, bereavement, unemployment, family history of abuse or neglect, family history of drug misuse, family history of alcohol misuse, family history of depression, abuse/neglect/non-accidental injury, neonatal health problems, missed immunisations, developmental delay, police involvement, other social services involvement, psychosexual problems, school problems, teenage pregnancy, work stress, young carer.

Data sources / Measurements

Clinical data are entered by general practitioners or other clinicians during routine consultations. Asthma, diabetes, dyspepsia and epilepsy were defined as present if either a clinical code (Read code) or prescription of a specific drug was recorded in the two years prior to the index date. The remaining exposure variables were defined as present if a clinical code was recorded in the two years prior to the index date, with the exception of developmental delay, early childhood emotional problems, missed immunisations, neonatal health problems and early/late puberty which were defined as present if a clinical code was *ever recorded* prior to the index date.

Smokers were defined as patients with any record indicating smoking in the two years prior to the index date. Patients who did not have a smoking status recorded remained in the analysis but had their smoking status categorised as missing. It is thought that this group might be predominantly non-smokers who had not been asked their smoking status, or patients who did not regularly visit their GP.

A count of the number of GP consultations in the year prior to the index date was made. For patients who had less than one year of registration (between six months and one year) their consultations over a six month period were counted and doubled. This was a continuous variable and therefore model estimates represent a linear relationship between the number of consultations and the probability of depression.

Study size

Survey data indicate that 2.2% of young people aged between 16-24 years experienced an episode of depression in the past week.[44] In 2009 there were 6,570,800 young people aged 15-24 in the UK, we would therefore expect approximately 144,500 to experience an episode of depression. As the THIN dataset is a broadly representative sample of approximately 6% of the UK population we would expect 8,670 cases of depression in our dataset. This is sufficient to investigate all conceivable predictor variables.[45]

Statistical methods

Because some predictors had been unstable over time in a previous similar analysis, an initial analysis was carried out to determine whether the relationships between the variables and depression were stable over time. Univariable odds ratios were calculated for each year of diagnosis and visualised using run charts: a systematic change in odds ratio over time would lead to the variable's exclusion. No exposure variables were excluded as a result of this exercise. In addition, frequency counts of each exposure variable were produced and variables with too few events (<0.02% of total sample size) were eliminated from the set of potential predictors. Eight variables were excluded because they were infrequently recorded in the dataset: sleep (too much), divorce, unemployment, teenage pregnancy, family history of abuse or neglect, family history of drug misuse, family history of alcohol misuse and family history of depression. Ethnicity was not included in the model predictors due to the amount of missing data (63.2% of patients had no ethnicity recorded). Patients who did not have Townsend quintile recorded were excluded from the analysis.

Two-thirds of practices were randomly allocated to be a development dataset (used for model development, selection of variables and estimation of regression coefficients) and the remaining third of practices allocated to a validation dataset (to test discrimination ability of model).

The statistical model was developed by entering all potential predictor variables into a backward stepwise conditional logistic regression model, with the significance level for removal of predictors set at 0.01 (lower than 0.05 due to large sample size). This method has been used to develop prediction models for anxiety and for depression in adults.[32, 33] Separate models were developed for two age groups within each gender: 15-19 years and 20-24 years. The final set of potential predictors available to all models included: Townsend quintile, smoking status, anxiety, low mood, tiredness, loss of enjoyment, sleep disorder (too little), eating disorders, weight loss, weight gain, bed wetting, self-harm, headache, dyspepsia, dysmenorrhea, back pain (with and without specific characteristics), ill-defined conditions, other somatic symptoms, diabetes, epilepsy, asthma, skin problems, childhood emotional problems, homelessness, bereavement, abuse/neglect/non-accidental injury, neonatal health problems, missed immunisations, developmental delay, police involvement, other social services involvement, psychosexual problems, school problems, work stress, young carer, OCD, PTSD, alcohol misuse, drug misuse, abdominal pain, excessive sweating, early/late puberty and number of consultations in the year prior to the index date.

As a sensitivity analysis, each model was developed, omitting the following variables from the set of potential predictors as these symptoms may have indicated that the GP was already considering depression as a possible diagnosis: anxiety, bereavement, low mood, self-harm, OCD and PTSD.

To investigate the discrimination characteristics of the final models, individuals in the validation dataset were allocated a score equal to their multivariable odds ratio and receiver operating characteristic (ROC) curves constructed for each model.

Additional analyses were performed to investigate whether the number of risk factors increased the risk of depression. A count of the number of risk factors in the following four groups was included in the model: symptoms of depression; somatic symptoms; co-morbidities; family and social factors.

All analyses were performed using Stata (version 12). Clinical code and drug lists are available from the authors on request.

Results

A total 98,562 cases and 281,248 controls were selected from 564 general practices (Figure 1). Most of the cases were female (67.1%) and diagnosed with depression between the ages of 20 and 24. Demographic and frequency of occurrence of exposure variables are shown in Table 1. Although the original aim was to match three controls to each case, this was not possible for every case, thus the case to control ratio achieved was 1:2.85. Where possible controls were matched to cases of the same age in years, and this was possible for 98.6% of controls, the remainder being matched with a control closest in age up to 3 years older/younger.

The development dataset consisted of 67,321 cases and 192,135 controls, and the validation dataset had 31,241 cases and 89,113 controls. Stepwise conditional logistic regression modelling was carried with 42 potential predictors plus Townsend quintile and smoking status; the final model for each dataset is presented in Table 2. Excluding specific symptoms which might be indicative of depression from the variable selection process did not result in any additional variables coming in to the final models.

Figure 2 shows receiver operating characteristic curves for each model, produced using the validation dataset. The area under the curve was similar for all four models; males aged 15 – 18: 0.71 (95% CI 0.70 to 0.73), males aged 19 – 24: 0.72 (95% CI 0.71 to 0.72), females aged 15 – 18: 0.72 (95% CI 0.71 to 0.73), females aged 19 – 24: 0.70 (95% CI 0.69 to 0.70).

Sensitivity analyses, where the model was developed omitting potential predictors which might be indicative of early signs of depression, resulted in only minor differences in the variables included and estimates of effect. Adding the number of potential risk factors did give a significant effect for some of the risk factor groups, although this did not result in a significant improvement to the model fit.

Discussion

Findings

Our analysis of a large dataset of electronic primary care records identified a number of socioeconomic and symptomatic predictors of a diagnosis of depression in young males and females aged 15 to 18 years and 19 to 24 years. Whilst the multivariable models derived for males had better discrimination characteristics than those derived for females, a number of predictors were common to all models. These included Townsend quintile, smoking status, symptoms of depression (anxiety, low mood, tiredness, too little sleep and self-harm); somatic symptoms (headache, back pain, dyspepsia, frequent consultation); life events (bereavement, indicators of abuse or neglect) and other psychological conditions (obsessive compulsive disorder). The strongest predictors were symptoms of depression and other psychological conditions. School problems (bullying, school refusal and truancy) and social services involvement were more prominent predictors in males than females aged 15 to 18 years, whereas work stress was only a predictor in females aged 19 to 24 years.

It is possible to derive an estimate of the probability of depression in the next year using Bayes Theorem and assuming no interaction between age and predictors. The annual incidence of depression (D) is an estimate of the prior probability of depression and the odds ratio is an estimate of the positive likelihood ratio (LR). The post-test probability of depression is given by (D/(1-D)×LR)/(1-(D/(1-D)×LR). More than half of females aged 18 to 23 and more than one in five males aged 19 to 24 with an odds ratio of 10 will be diagnosed with depression within a year (Table 3). This is consistent with having two or three predictors of depression.

Strengths

The recording of data in this large primary care dataset is reflective of usual practice in primary care. This means that a prediction model makes use of readily available data and is therefore in a general practice setting. This distinguishes it from previous prediction models, which include specific data items collected in the context of a research project.[32,33] The size of the dataset has allowed an extensive number of potential predictor variables to be included in the analysis. As with other prediction models using primary care records data, the added complexity of including multiple predictors can be mitigated by integrating the prediction tool into database software. [35]

Limitations

The main limitations of the model are the accuracy and completeness of records. If depression in some patients is never diagnosed this may weaken the associations between predictors and outcomes. Depression is not always diagnosed and some predictors of depression are infrequently elicited or recorded, particularly family and relevant social histories. This could be addressed by testing the model prospectively on a cohort of young adults. Prescription of an antidepressant drug was taken to indicate a diagnosis of depression. This may misclassify some patients as depressed, in particular those with obsessive compulsive disorder who may be treated with antidepressants.

Factors such as divorce, unemployment, teenage pregnancy, family history of abuse or neglect, family history of drug misuse, family history of alcohol misuse and family history of depression were excluded from our study due to low levels of recording. Further analysis might group these variables using factor analysis or latent class analysis to identify clusters of predictors. Incomplete recording of ethnicity, also meant that it was not included in the analysis. Not all presenting symptoms are recorded for each consultation and recording of symptoms may be more likely when a diagnosis of depression is being considered, exaggerating the association between symptoms and diagnosis.

Comparison to existing research

Predictors of depression in adults consistently include previous history of depression, family history of psychological difficulties, physical health problems, mental health problems (assessed by Short Form 12) and difficulties in paid or unpaid work.[33] Our analysis identified a similar range of factors reflecting the work and school environment, family circumstances and personal health problems. Specific factors in young men aged 15-18 years included school problems (truancy, bullying, school refusal) and social services involvement. These findings support existing research which has shown high levels of depression (90%) in young people presenting with mixed school refusal (both anxious school refusers and truants)[46], those who experience bullying,[47, 48] in particular, those both participating in and experiencing bullying[49] and those experiencing unpredictable, chaotic or abusive interpersonal relationships.[50, 51] Three quarters of these mixed school refusers also had a parent with a mental health problem, which is also a risk factor for depression in young people. [52]

Future research

This case-control study is a promising first step in to deriving a predictive model to assist primary care clinicians to improve their clinical awareness and diagnosis of depression in young people. A retrospective cohort design would allow a direct estimation of risk of depression related to symptoms and other patient characteristics.

References

[1] Department of Health (2004) Mental health of children and young people in Great Britain. Office for National Statistics

[2] Costello EJ, Erkanli A, Angold A: Is there an epidemic of child or adolescent depression? J Child Psychol Psychiatry 2006, 47:1263–1271.

[3] Garrison CZ, Addy CL, Jackson KL, McKeown RE, Waller JL Major depressive disorder and dysthymia in young adolescents *Am J Epidemiol* 1992;135792- 802

[4] Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA Adolescent psychopathology, I: prevalence and incidence of depression and other DSM-III-R disorders in high school students J Abnorm Psychol 1993;102133-144.

[5] Shaffer DFisher PDulcan MK et al. The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): description, acceptability, prevalence rates, and performance in the MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study J Am Acad Child Adolesc Psychiatry 1996;35865- 877.

[6] Whitaker A, Johnson J, Shaffer D et al. Uncommon troubles in young people: prevalence estimates of selected psychiatric disorders in a non-referred adolescent population Arch Gen Psychiatry 1990;47487- 496.

[7] Lewinsohn, P. M., Roberts, R. E. et al (1994a). 'Adolescent psychopathology: II. Psychosocial risk factors for depression', J Abnorm Psychol, 103 (2), 302-315.

[8] Harrington, R., Fudge, H., Rutter, M., *et al* (1990) Adult outcomes of child and adolescent depression. 1: Psychiatric status. *Archives of General Psychiatry*, **47**, 465–47

[9] Lewinsohn, P. M., Rhode, P., Seeley, J. R., *et al* (2000) Natural course of adolescent major depressive disorder in a community sample: predictors of recurrence in young adults. *American Journal of Psychiatry*, **157**, 1584–1591.

[10] Patton GC, Coffey C, Romaniuk H, Mackinnon A, Carlin JB, Degenhardt L, Olsson CA, Moran P. The prognosis of common mental disorders in adolescents: a 14-year prospective cohort study. Lancet. 2014 Apr 19;383(9926):1404-11. doi: 10.1016/S0140-6736(13)62116-9. Epub 2014 Jan 16.

[11] Royal College of Paediatrics and Child Health. Making the UK's child health outcomes comparable to the best in the world. A vision for 2015 http://www.rcpch.ac.uk/system/files/protected/news/RCPCH%20Child%20Health%20Manifesto%20WEB.pdf [Last accessed 2nd Septemer 2015]

[12] Knapp M, McDaid D & Parsonage M. Mental health promotion and mental illness prevention: The economic case. Centre for Mental Health & Longdon School for Economics, Department of Health 2011.

[13] Allen G. Early Intervention : The Next Steps. HM Government 2011.

[14] Department of Health (2006) Children and young people in mind: the final report of the National CAMHS Review

[15] Zachrisson HD, Rödje K, Mykletun A. Utilization of health services in relation to mental health problems in adolescents: a population based survey. *BMC Public Health* 2006; 6: 34-40.
[16] Wisdom JP, Clarke GN, Green CA. What teens want: barriers to seeking care for depression. Adm Policy Ment Health 2006; 33: 133-145
[17] Hirschfeld RM, Keller MB, Panico S, et al. The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *JAMA*. 1997;277(4):333-340.
[18] Williams SB, O'Connor EA, Eder M, Whitlock EP. Screening for child and adolescent depression in primary care settings: a systematic evidence review for the US Preventive Services Task Force. Pediatrics. 2009 Apr; 123(4):e716-35.
[19] Beck AT, Guth D, Steer RA, et al. Screening for major depression disorders in medical inpatients with the Beck depression inventory for primary care. Behav Res Ther 1997;35:785–91
[20] Spitzer RL, Johnson JG. The Patient Health Questionnaire, Adolescent Version Biometrics Research Unit, New York State Psychiatric Institute (1995)
[21] Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337–45
[22] Sharp LK & Lipsky MS. Screening for Depression Across the Lifespan:A Review of Measures for Use in Primary Care Settings. American Family Physician. 2002; 66 (6):1001-1008
[23] Newton S, Docter S, Reddin E, Merlin T, Hiller JE. (2010) Depression in adolescents and young adults. Evidence Review. Adelaide SA: beyondblue.
[24] Hadjiyannakis, K. K. (2003) In Dissertation Abstracts International: Section B: The Sciences and Engineering, Vol. 64 ProQuest Information & Learning, US, pp. 965
[25] Roane BM & Taylor DJ (2008) Adolescent insomnia as a risk factor for early adult depression and substance abuse. Sleep 31(10): 1351–56.
[26] Bohman H, Jonsson U, von Knorring A, von Knorring L,Paaran A & Olsson G. (2010). Somatic Symptoms as a marker for severity in adolescent depression. Acta Paediatrica 99, pp. 1724–1730
[27] Keenan-Miller D, Hammen CL, Brennan PA. Health outcomes related to early adolescent depression. J Adolesc Health. 2007 Sep;41(3):256-62.
[28] Haarasilta LM, Marttunen MJ, Kaprio JA, Aro HM. Correlates of depression in a representative nationwide sample of adolescents (15-19 years) and young adults (20-24 years). Eur J Public Health. 2004 Sep;14(3):280-5.
[29] Steuber TL, Danner F. Adolescent smoking and depression: which comes first? Addict Behav. 2006 Jan;31(1):133-6.
[30] Fergusson DM, Goodwin RD, Horwood LJ. Major depression and cigarette smoking: results of a 21-year longitudinal study. Psychol Med. 2003 Nov;33(8):1357-67.
[31] Barbeau E, Krieger N & Soobader M. Working class matters: Socioeconomic disadvantage, race/ethnicity, gender and smoking in NHIS 2000. American Journal of Public Health 94 (2) 269-278.
[32] King M, Bottomley C, Bellón-Saameño JA, Torres-Gonzalez F, Švab I, Rifel J, Maaroos HI, Aluoja A, Geerlings MI, Xavier M, Carraça I, Vicente B, Saldivia S, Nazareth I. An international risk prediction algorithm for the onset of generalized anxiety and panic syndromes in general practice attendees: predictA. Psychol Med. 2011 Aug;41(8):1625-39. doi: 10.1017/S0033291710002400. Epub 2011 Jan 6.
[33] King M, Walker C, Levy G, Bottomley C, Royston P, Weich S, Bellón-Saameño JA, Moreno B, Svab I, Rotar D, Rifel J, Maaroos HI, Aluoja A, Kalda R, Neeleman J, Geerlings MI, Xavier M, Carraça I, Gonçalves-Pereira M, Vicente B, Saldivia S, Melipillan R, Torres-Gonzalez F, Nazareth I. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. Arch Gen Psychiatry. 2008 Dec;65(12):1368-76. doi: 10.1001/archpsyc.65.12.1368.
[34] Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract. 2013 Jan;63(606):e11-21. doi: 10.3399/bjgp13X660733.
[35] Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ. 2009 Mar 17;338:b880. doi: 10.1136/bmj.b880.
[36] Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. BMJ Open. 2013 Aug 19;3(8):e003482. doi: 10.1136/bmjopen-2013-003482.
[37] Cegedim Strategic Data http://csdmruk.cegedim.com/our-data/our-data.shtml [Accessed 25th March 2014]
[38] Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Informatics in Primary Care 2011;19(4):251-5.
[39] Cegedim Strategic Data http://csdmruk.cegedim.com/our-data/ethics.shtml [Accessed 25th March 2014]

[40] Maguire A1, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. Pharmacoepidemiol Drug Saf. 2009 Jan;18(1):76-83. doi: 10.1002/pds.1688.
[41] Blakemore SJ, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. J Child Psychol Psychiatry 2006 Mar-Apr;47(3-4):296-312.
[42] Health & Social Care Information Centre – Read Codes http://systems.hscic.gov.uk/data/uktc/readcodes [Accessed 28th March 2014]
[43] British National Formulary http://www.bnf.org/bnf/index.htm [Accessed 28th March 2014]
[44] Health and Social Care Information Centre Adult Psychiatric Morbidity in England - 2007, Results of a household survey [NS] Publication date: January 27, 2009 http://www.hscic.gov.uk/pubs/psychiatricmorbidity07 [Accessed 11th December 2014]
[45] Steyerberg EW Clinical prediction models: a practical approach to development, validation and updating. Springer 2010.
[46] Egger H, Costello EJ & Angold A. School Refusal and Psychiatric Disorders: A Community Study. J. Am. Acad. Child Adolesc. Psychiatry, 2003; 42:7: 797-807.
[47] Goodyer IM, Wright C, Altham P. The friendships and recent life events of anxious and depressed school-age children. British Journal of Psychiatry. 1990;156:689–698
[48] Goodyer IM, Herbert J, Tamplin A, et al. Recent life events, cortisol, dehydroepiandrosterone and the onset of major depression in high-risk adolescents. British Journal of Psychiatry. 2000;177:499–504
[49] Kaltiala-Heino R, Rimpela M, Rantanen P & Rimpela A. Bullying at school – An indicator of adolescents at risk for mental disorders. Journal of Adolescence, 2000: 23: 661-674.
[50] Jaffee SR, Moffitt TE, Caspi A, et al. Differences in early childhood risk factors for juvenile-onset and adult-onset depression. Archives of General Psychiatry. 2002;59:215–222
[51] Hammen C, Shih JH, Brennan PA. Intergenerational transmission of depression: test of an interpersonal stress model in a community sample. Journal of Consulting and Clinical Psychology. 2004;72:511–522.
[52] Hammen C, Brennan PA. Severity, chronicity, and timing of maternal depression and risk for adolescent offspring diagnoses in a community sample. Archives of General Psychiatry. 2003;60:253–258

Tables and Figures

Table 1: Characteristics of cases and controls

Variable	Cases, n (%) n = 98,562	Controls, n (%) n = 281,248
Male	32,470 (32.9%)	96,444 (34.4%)
Age at index date (years)		
Under 15	0	87 (0.03%)
15-19	36,796 (37.3%)	105,935 (37.7%)
20-24	61,766 (62.7%)	174,612 (62.1%)
Over 24	0	614 (0.2%)
Ethnicity		
White	34,488 (35.0%)	89,980 (32.0%)
Black	709 (0.7%)	3,075 (1.1%)
Asian	1,223 (1.2%)	5,691 (2.0%)
Mixed	484 (0.5%)	1,436 (0.5%)
Chinese	45 (0.1%)	881 (0.3%)
Other	360 (0.4%)	1,584 (0.6%)
Missing	61,253 (62.2%)	178,601 (63.5%)
Townsend		
1 (Least deprived)	16,530 (16.8%)	56,911 (20.2%)
2	15,713 (15.9%)	50,318 (17.9%)
3	19,747 (20.0%)	56,217 (20.0%)
4	23,233 (23.6%)	60,141 (21.4%)
5	19,721 (20.0%)	46,073 (16.4%)
Missing	3,618 (3.7%)	11,588 (4.1%)
Smoking status		
Smoker	37,571 (38.1%)	68,066 (24.2%)
Ex/non- smoker	45,918 (46.6%)	148,679 (52.9%)
Missing	15,073 (15.3%)	64,503 (22.9%)
Symptoms of depression		
Anxiety	4,919 (5.0%)	3,107 (1.1%)
Low mood	5,814 (5.9%)	2,362 (0.8%)
Tiredness	2,901 (2.9%)	3,272 (1.2%)
Loss of enjoyment	130 (0.1%)	142 (0.1%)
Sleep disorder (too little)	886 (0.9%)	619 (0.2%)
Sleep disorder (too much)	32 (<0.1%)	29 (<0.1%)
Eating disorders	909 (0.9%)	698 (0.3%)
Weight loss	884 (0.9%)	993 (0.4%)
Weight gain	175(0.2%)	268 (0.1%)
Excessive sweating	583 (0.6%)	1,048 (0.4%)
Bed wetting	120 (0.1%)	203 (0.1%)
Self-harm	1,478 (1.5%)	814 (0.3%)
Somatic symptoms		
Headache	14,430 (14.6%)	18,823 (6.7%)
Dyspepsia	11,989 (12.2%)	15,090 (5.4%)
Dysmenorrhea	3,069 (3.1%)	5,669 (2.0%)
Abdominal pain	7,504 (7.6%)	10,410 (3.7%)
Early/late puberty	228 (0.2%)	548 (0.2%)
Back pain: with specific characteristics	1,356 (1.4%)	1,834 (0.7%)
Back pain: without specific characteristics	10,846 (11.0%)	16,175 (5.8%)
Ill-defined conditions	499 (0.5%)	775 (0.3%)
Other somatic symptoms	112 (0.1%)	81 (<0.1%)
Co-morbidities		
Diabetes	1,252 (1.3%)	1,725 (0.6%)
Epilepsy	1,235 (1.3%)	2,347 (0.8%)
Asthma	15,637 (15.9%)	29,982 (10.7%)
Skin problems.	14,280 (14.5%)	33,186 (11.8%)
Drug and alcohol use		

Variable	Cases, n (%) n = 98,562	Controls, n (%) n = 281,248
Alcohol misuse	753 (0.8%)	876 (0.3%)
Drug misuse	998 (1.0%)	804 (0.3%)
Family and social factors		
Childhood emotional problems	60 (<0.1%)	66 (<0.1%)
Divorce	12 (<0.1%)	12 (<0.1%)
Homelessness	83 (0.1%)	84 (<0.1%)
Bereavement	1,171 (1.2%)	852 (0.3%)
Unemployment	9 (<0.1%)	6 (<0.1%)
Family history of abuse or neglect	9 (<0.1%)	12 (<0.1%)
Family history of alcohol misuse	2 (<0.1%)	2 (<0.1%)
Family history of drug misuse	5 (<0.1%)	4 (<0.1%)
Family history of depression	0	0
Abuse/neglect/non-accidental injury	1,829 (1.9%)	2,055 (0.7%)
Neonatal health problems	8,641 (8.8%)	21,429 (7.6%)
Missed immunisations	662 (0.7%)	1,592 (0.6%)
Developmental delay	2,253 (2.3%)	5,410 (1.9%)
Police involvement	45 (0.1%)	53 (<0.1%)
Other social services involvement	80 (0.1%)	115 (<0.1%)
Psychosexual problems	296 (0.3%)	313 (0.1%)
School problems	334 (0.3%)	268 (0.1%)
Teenage pregnancy	15 (<0.1%)	18 (<0.1%)
Work stress	75 (0.1%)	54 (<0.1%)
Young carer	83 (0.1%)	176 (0.1%)
Other psychological conditions		
Post-traumatic stress disorder	188 (0.2%)	113 (<0.1%)
Obsessive compulsive disorder	436 (0.4%)	193 (0.1%)

Table 2: Results of multivariable conditional logistic regression analysis for depression prediction

	Males						Females					
	15 – 18y (4,702 cases / 14,074 controls)			19 – 24y (17,526 cases / 51,907 controls)			15 – 18y (11,857 cases / 34,315 controls)			19 – 24y (33,236 cases/91,839 contro		
	OR	**(95% CI)**	**p-value**	**OR**	**(95% CI)**	**p-value**	**OR**	**(95% CI)**	**p-value**	**OR**	**(95% CI)**	**p-v**
Townsend quintile (ref=1 (least deprived)												
2	0.98	(0.87 to 1.11)	0.770	1.09	((1.02 to 1.17)	0.010	1.08	(1.00 to 1.17)	0.041	1.07	(1.02 to 1.13)	
3	1.16	(1.03 to 1.31)	0.017	1.17	(1.10 to 1.25)	<0.001	1.23	(1.14 to 1.33)	<0.001	1.13	(1.07 to 1.19)	<
4	1.23	(1.09 to 1.40)	0.001	1.33	(1.24 to 1.42)	<0.001	1.28	(1.18 to 1.38)	<0.001	1.22	(1.16 to 1.28)	<
5 (most deprived)	1.56	(1.35 to 1.80)	<0.001	1.47	91.36 to 1.57)	<0.001	1.35	(1.23 to 1.47)	<0.001	1.29	(1.22 to 1.36)	<
Smoking status (ref=ex/non-smoker)												
Smoker	1.88	(1.66 to 2.11)	<0.001	1.81	(1.73 to 1.89)	<0.001	1.35	(1.27 to 1.44)	<0.001	1.56	(1.51 to 1.61)	<
Missing	0.87	(0.78 to 0.96)	0.005	1.00	(0.95 to 1.07)	0.751	0.83	(0.77 to 0.88)	<0.001	0.90	(0.85 to 0.96)	<
Symptoms of depression												
Anxiety	6.03	(4.49 to 8.09)	<0.001	5.41	(4.69 to 6.24)	<0.001	3.26	(2.78 to 3.82)	<0.001	2.86	(2.63 to 3.11)	<
Low mood	10.25	(7.38 to 14.23)	<0.001	10.40	(8.63 to 12.52)	<0.001	5.49	(4.79 to 6.31)	<0.001	4.67	(4.27 to 5.11)	<
Tiredness	3.10	(2.03 to 4.73)	<0.001	2.24	(1.84 to 2.73)	<0.001	2.02	(1.72 to 2.37)	<0.001	1.78	(1.63 to 1.95)	<
Loss of enjoyment										1.73	(1.22 to 2.46)	
Sleep disorder (too little)	4.27	(2.40 to 7.62)	<0.001	2.09	(1.57 to 2.77)	<0.001	2.51	(1.81 to 3.48)	<0.001	2.05	(1.67 to 2.52)	<
Eating disorders				2.13	(1.32 to 3.42)	0.002	2.30	(1.83 to 2.89)	<0.001	2.72	(2.26 to 3.28)	<
Weight loss				1.84	(1.40 to 2.42)	<0.001				1.58	(1.32 to 1.88)	<
Excessive sweating										1.29	(1.07 to 1.56)	
Bed wetting	2.98	(1.56 to 5.70)	<0.001									
Self-harm	8.22	(4.92 to 13.73)	<0.001	4.77	(3.57 to 6.37)	<0.001	3.38	(2.81 to 4.06)	<0.001	3.33	(2.68 to 4.13)	<
Somatic symptoms												
Headache	2.30	(1.99 to 2.67)	<0.001	2.14	(1.97 to 2.33)	<0.001	1.75	(1.63 to 1.88)	<0.001	1.71	(1.63 to 1.78)	<
Dyspepsia	1.74	(1.44 to 2.11)	<0.001	1.41	(1.30 to 1.53)	<0.001	1.50	(1.37 to 1.64)	<0.001	1.39	(1.33 to 1.46)	<
Dysmenorrhea										1.20	(1.09 to 1.31)	<
Abdominal pain				1.48	(1.31 to 1.66)	<0.001	1.32	(1.19 to 1.46)	<0.001	1.21	(1.14 to 1.28)	<
Back pain: with specific characteristics										1.20	(1.05 to 1.37)	
Back pain: without specific characteristics	1.47	(1.23 to 1.75)	<0.001	1.38	(1.28 to 1.48)	<0.001	1.29	(1.17 to 1.41)	<0.001	1.36	(1.30 to 1.43)	<
Number of consultations in year	1.17	(1.15 to 1.19)	<0.001	1.14	(1.13 to 1.15)	<0.001	1.11	(1.10 to 1.12)	<0.001	1.08	(1.08 to 1.09)	<
Co-morbidities												
Diabetes				1.95	(1.59 to 2.40)	<0.001				1.41	(1.23 to 1.62)	<
Epilepsy				0.77	(0.64 to 0.94)	0.009						
Asthma										1.16	(1.11 to 1.21)	<
Drug and alcohol use												
Alcohol misuse				1.68	(1.34 to 2.11)	<0.001				1.46	(1.15 to 1.85)	<

	Males						Females					
	15 – 18y (4,702 cases / 14,074 controls)			19 – 24y (17,526 cases / 51,907 controls)			15 – 18y (11,857 cases / 34,315 controls)			19 – 24y (33,236 cases/91,839 contro		
	OR	(95% CI)	p-value	OR	(95% CI)	p-value	OR	(95% CI)	p-value	OR	(95% CI)	p-v
Drug misuse	2.51	(1.43 to 4.37)	<0.001	2.08	(1.73 to 2.51)	<0.001				2.09	(1.63 to 2.68)	<
Family and social factors												
Bereavement	2.93	(1.59 to 5.38)	0.001	3.63	(2.77 to 4.74)	<0.001	2.24	(1.66 to 3.01)	<0.001	3.23	(2.74 to 3.81)	<
Abuse/neglect/non-accidental injury	1.64	(1.16 to 2.30)	<0.001	1.77	(1.49 to 2.10)	<0.001	1.57	(1.30 to 1.89)	<0.001	1.65	(1.41 to 1.92)	<
Neonatal health problems				1.15	(1.08 to 1.24)	<0.001				1.12	(1.06 to 1.19)	<
Developmental delay				1.17	(1.04 to 1.32)	<0.001						
Other social services involvement	4.89	(1.79 to 13.35)	0.002									
Psychosexual problems				2.12	(1.66 to 2.73)	<0.001						
School problems	5.84	(3.51 to 9.71)	<0.001				2.04	(1.52 to 2.73)	<0.001			
Work stress										3.05	(1.77 to 5.24)	<
Other psychological conditions												
Post-traumatic stress disorder				4.07	(2.30 to 7.21)	<0.001	3.33	(1.66 to 6.70)	0.001	2.53	(1.41 to 4.53)	
Obsessive compulsive disorder	13.98	(7.07 to 27.66)	<0.001	9.89	(5.93 to 16.51)	<0.001	8.57	(5.24 to 14.03)	<0.001	3.45	(2.39 to 4.97)	<

Table 3: Prior and posterior probability of a diagnosis of depression in the next year if the multivariable odds ratio is 10

Age (years)	Annual incidence of depression per 1000 person years			
	Average incidence 2000 - 2012		Predicted if multivariable Odds Ratio is 10	
	Male	Female	Male	Female
15	3.5	9.6	37	108
16	4.8	14.5	51	173
17	8.0	25.1	87	346
18	12.2	35.0	141	569
19	16.6	40.7	204	738
20	18.1	40.6	226	733
21	18.1	39.6	225	701
22	18.6	36.3	234	604
23	17.7	32.8	220	514
24	17.3	30.8	214	465

Source: Depression incidence data from THIN 2000 to 2012

Figure 1: Flow diagram of case/control selection

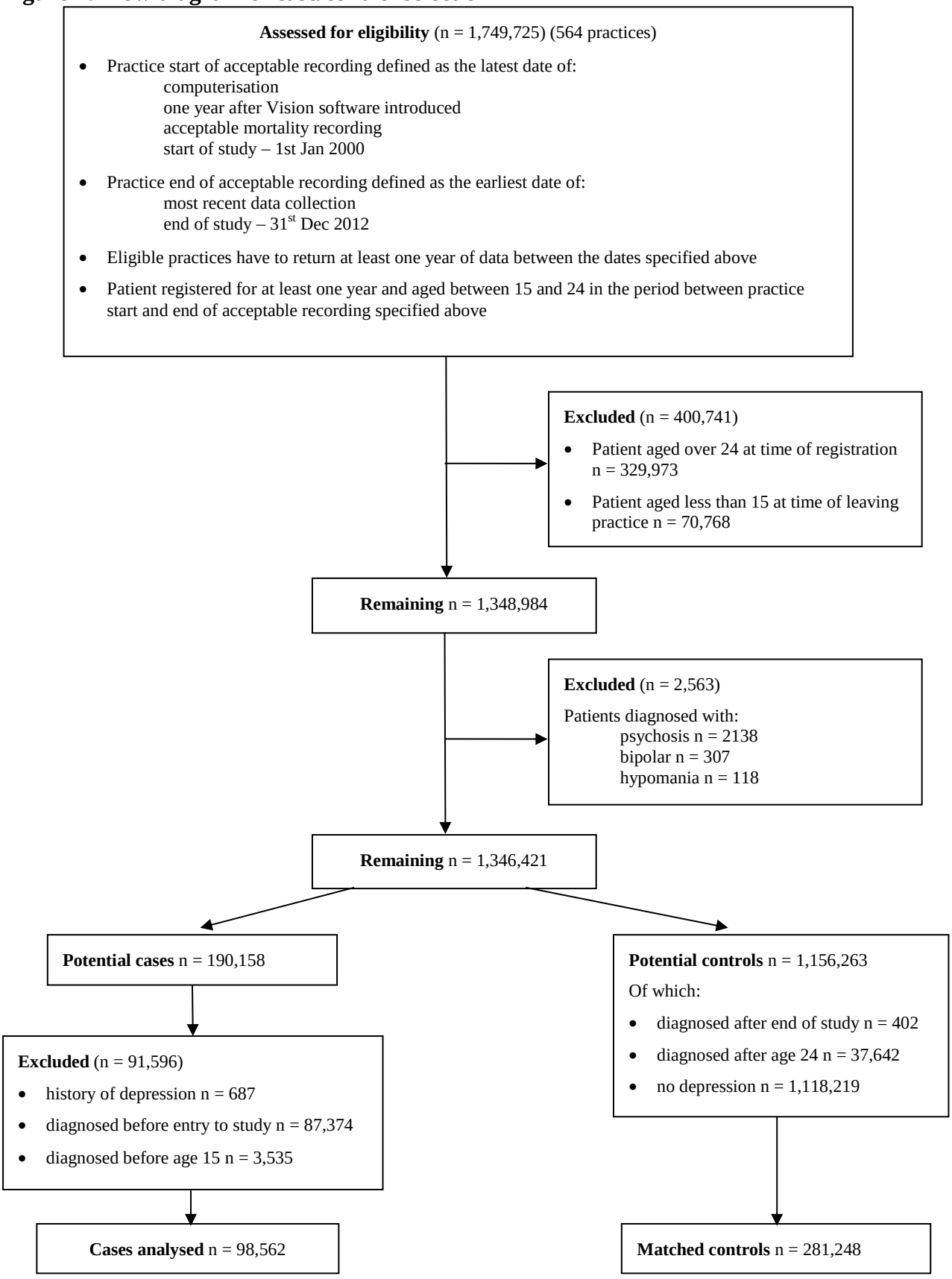
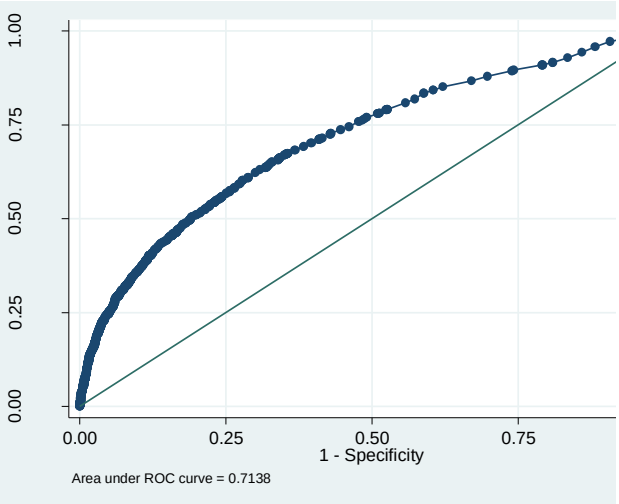
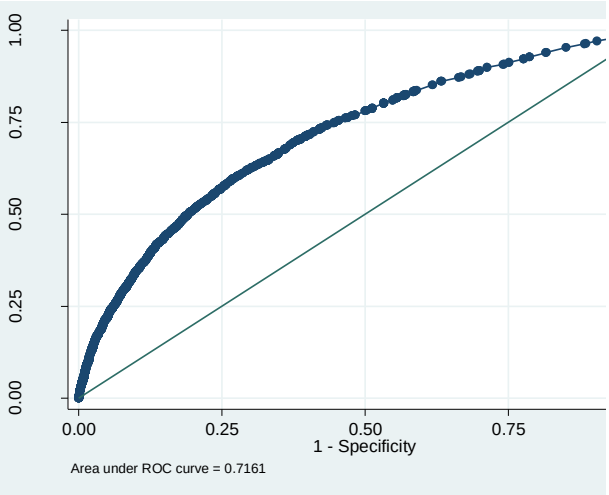


Figure 2: Receiver operating characteristics curves for conditional logistic models

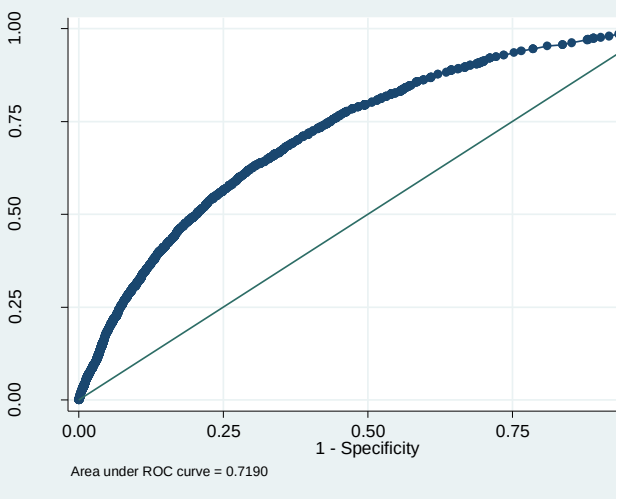
Males, 15 - 18y



Males, 19 - 24y



Females, 15 – 18y



Females 19 – 24y