PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A population-level prediction tool for the incidence of first episode psychosis: translational epidemiology based on cross-sectional data
AUTHORS	Kirkbride, James; Jackson, Daniel; Perez, Jesus; Fowler, David; Winton, Francis; Coid, Jeremy; Murray, Robin; Jones, Peter

VERSION 1 - REVIEW

REVIEWER	Professor John McGrath Queensland Brain Institute University of Queensland Brisbane QLD 4072 Australia
REVIEW RETURNED	05-Sep-2012

GENERAL COMMENTS	This paper uses high quality, population-based incidence of schizophrenia data to 'map' small-area based estimates. Based on variables known to be associated with risk of schizophrenia (e.g. age, sex, ethnicity, SES variables etc), several models were presented in order to explore incidence in an independent population-based incidence study (replication dataset). The best model was found to include age, sex, age by sex interaction and population density. In addition, the models have been used to underpin an innovative (and attractive) web-based mapping. The study will be of interest to service planner (where first episode services would be needed, predicted case loads etc) and epidemiology (who rarely are able to inspect small-area level data for schizophrenia). The paper is well written, which was difficult to do in light of the complexity of some of the statistics and because of some fine-grained issues with the underlying data. I cannot comment on all details of the statistical methods. Comments: 1. The paper does not discuss issues related to internal migration prior to the onset of a psychotic disorder. If these data were available from any of the studies, it might be worth commenting on. 2. The SEPEA study excluded 44 participants (8% of those who entered the service), because they were not psychotic (a false negative, from one perspective). But, usually these individuals are distressed and needing care (assessment and triage at least). If the software is pitched to service planners, perhaps another adjustment should be made to include the extra initial work load related to these cases. 3. It was interesting to see which variables were best able to predict the counts from the SEPEA study. While I appreciate that the main models may not be statistically different (leaving aside the 'straw man' MH-PIG estimates), could the authors include text that population density alone seems to capture more variance compared

	to models based on overall IMD or income, employment or deprivation. Can we make any strong inferences about this? 4. In a similar fashion, I think the authors should include text noting population density and ethnicity (and maybe age and sex also) act as proxy measures for a complex socially-patterned matrix of risk factors that would underlie the incidence of schizophrenia (i.e. these variables are not themselves the causal agent). Should this go into the disclaimer as well? 5. Concerning the mapping to the rest of England and Wales, I am not sure how many of them would have well-publicized youth mental health services. Better access and mental health literacy will facilitate more referrals and thus higher incidence rates. These factors may not be evenly distributed across the nation. Again, this does not invalidate the estimates, but if there are future mimatches between the PsyMaptic prediction and later observed counts, this may simply reflect pathway to care issues.

REVIEWER	Max Birchwood Professor of Youth Mental Health School of Psychology, University of Birmingham and Clinical Director, YouthSpace, birmingham and Solihull Mental Health Foundation Trust, Birmingham, UK.
REVIEW RETURNED	25-Sep-2012

THE STUDY	This paper describes a tool for predicting the incidence of first episode psychosis, based on known epidemiological gradients or gender, ethnicity, deprivation and population density. The prediction sample developed a model which was then validated by attempting to predict numbers of new cases managed by specialised early intervention services spread across East Anglia. Model 6 was chosen, providing the best prediction of new cases presenting to these teams over the study period. Two issues: 1. It is assumed that the number of cases in EIS is a satisfactory proxy for age-incidence , 16-35. It seems to me likely that there will be various filters operating before cases access EIS; for example, in the majority of cases EIS will receive cases via CMHTs who may not refer to EIS. This will certainly vary across England and Wales as the PIG was not clear about the position of EIS in the care pathway. Thus it may be that model 1 is the best predictor of actual incidence (ie. higher than cases in EIS), but model 6 provides the best prediction of those that find their way into EIS. A second problem is that the choice of model may be affected by the Socio-demographic profile of the validation sample. There is a sharp disparity between the ethnicity and deprivation profile of the prediction and validation samples in table 1; while this helps to support the validity of the prediction tool (ie. it predicts cases in a very different setting), is it not possible that in a very different setting (say ethnically diverse Birmingham), model 6 would fail to provide the best prediction? Would it not therefore be more prudent to assert that further validation samples will be required to improve accuracy of the tool?
GENERAL COMMENTS	I thought that this tool will be an invaluable resource for public health and commissioning and illustrates well that the evidence base for planning services in 2001 was hopelessly flawed! The paper will need an expert methodological review, but in light of earlier comments the limitations of case ascertainment as a proxy for incidence and its generalisability across EIS in England and Wales

	with different care pathways etc, it will need a strong health warning I would have thought.

VERSION 1 – AUTHOR RESPONSE

We are grateful to Profs. McGrath and Birchwood for their helpful comments on our manuscript. We hope that the changes we have made in response mean that the manuscript is now stronger.

First, we describe a change we have made on the basis of our own review. Second, then we respond to the referees. Please note that all page numbers referred to in our responses, below, refer to the page when "tracked changes" in Microsoft Word are switched on.

General revision

We have corrected our reading of the 2001 Mental Health Policy Implementation Guide's (MH-PIG) uniform incidence rate for young people experiencing a first episode of psychosis (FEP). The MH-PIG refers to 15 per 100,000 people per year in its description of anticipated caseloads citing "150 new cases per year" in a "population of around 1 million people" (p55). We now read this not as 1m people aged 14-35 years (the age range for early intervention services (EIS), but as 150 new cases per 1m total population (i.e. all ages), with 50 such services to be commissioned in England when the total population at the time was just over 50m. In 2001 the Census estimated the proportion of people aged 14-35 in England as 29.3% of the total population. Under this interpretation, the uniform rate for service provision (i.e. Model 7 in our paper) should not be 15 per 100,000 people per year aged 14-35 years but should be: [150/(1m*.293)]*100,000 = 51 per 100,000 people in this age range, per year.

This revision does not affect our main prediction models (models 1-6 all remain the same), but does alter the predictions based on the uniform rate (model 7) that are now corrected (see Table 2). Model fit diagnostics for Model 7 also changed slightly as a result (see Table 3), but did not affect the interpretation overall; the MH-PIG rate remained the poorest predictor of FEP. Under the MH-PIG uniform rate of 51 per 100,000 people per year (14-35 years) our data suggest that the Department of Health over-estimated demand in most services (particularly in rural communities), but under-estimated demand in urban areas (particularly London boroughs and parts of Birmingham). We have revised the text throughout our manuscript, accordingly (see for example, Introduction, p3, last paragraph). We have also revised our online PsyMaptic software to reflect this, which has been updated to version 0.4. We have added one map to Section 2a of PsyMaptic ("16-35 years"> "2a. Overall") to highlight regions where the rate predicted by PsyMaptic falls either significantly above or below the Department of Health's uniform rate. This map now supersedes our original Supplemental Figure 2 in the manuscript.

Reviewer 1

1. Internal migration: Unfortunately we did not have data on the residential histories of people in either the ÆSOP or ELFEP first episode studies. Such data would have allowed us to get traction on the extent of possible social "drift" of people into poorer or more urban communities, a particularly important issue for aetiological understanding. On the other hand, knowledge of the extent to which social drift operates on young people with FEP is not as critical for service planning at the point of service contact; service planning requires a useful estimate of the likely need in a given community, based on markers which have been shown to index variation in incidence (such as age, sex, ethnicity or population density). We suggest our models should primarily be used to inform services of likely demand for clinical care associated with first episode psychosis, independent of the extent to which the level of burden on a team is a cause or consequence of factors such as urbanicity. We agree with the reviewer that it would be important to understand the extent to which social drift operates for

aetiology and for prevention of disorder. We have not commented on this issue directly in the paper given the emphasis on service delivery rather than aetiology. Our models are better than the guidance in the MH-PIG and should therefore be helpful.

2. We agree on this point and have amended the text throughout the manuscript. We have sought to differentiate in our manuscript between the predictions our models are based on i.e. clinically relevant first episode psychosis and the broader range of psychopathology that EIS often see at referral and may have to provide triage to. Such baseline assessments are an important part of any service and were not included in either our model or the MH-PIG. We were unable to monitor the exact proportion of people referred to EIS in the SEPEA study but who were not taken on, so were unable to "adjust" our estimates. This is something that could be applied to future estimates if we can obtain reliable data on the proportion referred to EIS vs the proportion accepted. In the present paper we have acknowledge that our prediction models forecast the likely true incidence of disorder (within a prediction interval), but that any service commissioning needs to be based on a wider array of factors. These include this issue (false negatives & positives), as well as supply-side issues, such as how different EIS are organised (whether they also operate early detection services, whether they see all types of psychotic disorder or just selected disorders, pathway to care filters and relationships to other services – CAMHS, IAPT, CMHTs). All these factors will influence the level of resourcing needed to run EIS teams commensurate with the needs of local populations. For examples of where we have amended the text in this regard, please see: Abstract (p2) -last sentence of the "conclusions" section; Discussion (p14) – last sentence of first paragraph; Discussion (p15 – 2nd & 3rd paragraphs, p16 – first paragraph) - all tracked changes pertain to this point.

3. We have added a line to the results to state that Model 6 (with population density) reported the lowest RMSE error at EIS and LAD level of any model (see Results, p12, under "External model prediction & validation", sentence beginning "This model had the lowest …". RMSE is one measure of the fit of the predicted data to the observations, where lower error indicates better model fit. We are confident that this model was a better indicator of expected rates in East Anglia than any other model tested, though we acknowledge that with further validation data from other regions, other models may provide a better fit to the data. This point was highlighted by Reviewer 2 and we have advocated the need to validate our models in other regions using more observation data from EIS settings (see Discussion P17, final paragraph, sentence beginning "Ongoing monitoring…").

4. We agree with the reviewer that the variables included in our models are likely to be proxies for other more complex socially patterned risk factors. While aetiology was not the primary focus of this paper we have nonetheless acknowledged this fact (Discussion, p17, first paragraph, last sentence beginning "From an aetiological perspective…".

5. We have acknowledged this point in the first paragraph on p16 ("Acceptance rates to EIS may also be influenced by local community awareness of such services") but do not think that this is an important bias. The implementation of EIS was very effective with teams being established throughout the nation. We have no reason to believe mental health literacy will be systematically different between the centres in the study and those in the rest of the country.

Reviewer 2

1. This is an important point and we have adjusted our paper, including the title, to emphasise the point that EIS operate with differential filters, meaning that the true incidence in the population differs from the incepted incidence in EIS. (See for examples, our corrections to the Discussion on p15-16). We also took great care in our modelling and validation to ensure that the predicted cases from our model were comparable to the observed cases identified in the SEPEA study. All our models were based on population-based data, so provided different estimates of what we would consider the

expected incidence of disorder in the population, rather than a prediction of an incepted incidence. We also attempted to ensure that case ascertainment in the SEPEA study was as epidemiologically rigorous as possible, as were the two studies, ÆSOP and ELFEP, on which the models are based, given the constraint that we were limited to the identification of cases via EIS. Nevertheless, in East Anglia, the three EIS are the only major service point for young people with psychosis and work closely with other mental health services to ensure high completeness of potential cases are referred to services. It is possible that a few cases may languish in the wrong part of a mental health system, but our predictions may at least indicate this as a possible reason for observed cases being below the expected.

We used strict criteria for entry in the SEPEA study including the exclusion of subjects not resident in the catchment area at the time of presentation to services, exclusion of subjects transferred from another service and exclusion of accepted subjects who did not meet clinical criteria for FEP at 6 months after acceptance. The reviewer's point is concerned that the incepted incidence of FEP may be lower than the true incidence in the population. While we were unable to directly estimate such an effect in the SEPEA study we note that, if anything, services taking part in the study saw a very broad range of psychopathology requiring clinical attention, though a proportion of referrals were not accepted for psychosis by the teams, while a smaller proportion (8%) were accepted but later found not to have a primary diagnosis of psychotic disorder at 6 months after acceptance. This suggests that services, at least in East Anglia, are casting their net wide to identify potential cases of first episode psychosis. On the one hand this leads to elevation in the "false positive" rate, but on the other hand it might suggest that incepted rates of clinical disorder are closer to the true rate in the population than might otherwise be the case. We have no reason to believe that a model without population density (i.e. with only age, sex and ethnicity) would be better at predicting the true rate. Population density is an established indicator of psychosis risk, with elevated rates amongst those born, brought up and living in urban areas at the time of onset. We stand by our decision to propose model 6 for our prediction tool, but we have revised our paper to recognise that the decisions which influence funding of services will not (or should not) be based on clinical caseloads alone; the wider range of psychopathology presenting to services need to be considered, as does the filters which operate differentially throughout England and Wales (see p15-16 & also first para of p. 14). One way to test our models further would clearly be through validation in more regions, and we have recommended this in our discussion (see revision to the last paragraph on p17).

2. As per the reviewer's suggestion we have asserted that further validation of our model(s) will be required (see last paragraph, p17). We acknowledge that continued testing of our models in different settings and different populations will be important to develop and refine the model(s). Understanding where our models fall down and where they are reliable will be equally important. Our prediction models are strengthened by their ability to take empirical data from predominantly urban settings with rich variation in the incidence of psychotic disorder by age, sex and ethnicity, and extrapolate these to a markedly different population where they were validated. We have no reason to believe that our predictions will not perform well in other areas because we believe the data on which they are founded are robust for the age, sex and ethnic groupings we included in our study. One area for future development will be to assess the validity of our models in specific age, sex and ethnic groups rather than at EIS, LAD or regional levels alone. We encourage our predictions to be used with their 95% prediction intervals which reflect our confidence in the predictions – for some specific ethnic groups these will be imprecise in several regions where population numbers are also absolutely low. An implicit assumption of our models is that the excess or reduced risk of psychosis amongst people from a given ethnic group (relative to another group) in one region is representative of the risk experienced by people from that group in other regions. For example, our model assumes that the elevated risk of psychosis amongst the black Caribbean population in AESOP and ELFEP (RR: 6.0; 95% CI: 4.9, 7.3) is the same average elevation in risk amongst black Caribbean communities in Birmingham, Liverpool, Manchester, Cardiff or Cornwall. This assumption will also require testing. We

know for example that ethnic density conditions risk in ethnic minority groups and so modelling such relationships may be important in future studies. We have added a line to this effect in the paper (see Discussion, first paragraph p15), but we were unable to perform such modelling in the present prediction models because 2011 Census data is not yet available at the required neighbourhood level (electoral wards).

We are delighted that the reviewer thought that the tool will be an invaluable resource for public health and commissioning and that it illustrates well that the evidence base for planning services in 2001 was hopelessly flawed; we concur with these views.

VERSION 2 – REVIEW

REVIEWER	John McGrath Director, Queensland Centre for Mental Health Research University of Queensland Australia No COI.
REVIEW RETURNED	23-Nov-2012

GENERAL COMMENTS	The authors have updated the paper based on the reviewers suggestions and also an important issue that they detected after submission

REVIEWER	Thomas P. A. Debray PhD Student Julius Center for Health Sciences and Primary Care The Netherlands
REVIEW RETURNED	30-Nov-2012

GENERAL COMMENTS	1. In section Data Generation: "Data were stratified by age group, sex, ethnicity and LAD, such that each stratum represented the total count of FEP cases in a unique sociodemographic group for a given LAD, with a corresponding estimate of the population at-risk, treated as an offset in our models." How many strata did this approah yield, and how did they look like? Did the authors verify whether sufficient strata were available to allow cross-validation with K=10? It would be useful to present some summaries about the generated strata, possibly in the appendix. 2. The authors employ a cross-validation approach that relies on random sampling of strata into training and testing subsets. This approach bears strong resemblance to "internal-external cross-validation", originally proposed by Royston et al. The authors evaluate the performance in the test subsets by overall measures of fit, including a correlation coefficient and a Mean Squared Error. These measures indicate whether derived models provide accurate predictions over all strata, however, it would also be useful to evaluate how derived models perform in individual strata. Particularly, inspecting the calibration in particular strata may help to identify to which extent the overall model may generalize towards

	new strata that were not considered during model development. For instance, the performance between rural and urban settings or across different age groups is not assessed during internal validation, but may be useful during external validation as included age groups only covered 16-35 (vs. 16-64 in the prediction sample). In general, detailed insight into internal validation results may help to identify boundaries of model generalizability, which are particularly useful when extrapolating the prediction model to new settings. For instance, this information could guide "Extrapolation to the United Kingdom" by pin-pointing strata for which accurate performance may be expected, and strata for which model predictions are prone to bias. References: Royston, P., Parmar, MKB. and Sylvester, R. (2004). Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. Statistics in Medicine 23: 907-926. 3. The discussion states "This model also had good internal validity across the entire age range (16-64 years)." , but it is not clear to me if internal validity was good in overall (see previous item), or whether calibration was indeed accurate for individual test strata. 4. Tables: please explain abbreviation IRR

REVIEWER	Collins, Gary University of Oxford, Centre for Statistics in Medicine
REVIEW RETURNED	18-Dec-2012

GENERAL COMMENTS	A complex study to develop a prediction model for incidence of first episode psychosis. The paper appears generally well conducted and reported, but I have some minor points the authors may wish to look at. Evaluating predictive ability of such a complex modelling approach is non-trivial and the authors appear to have done an adequate job. They have also responded adequately to all the initial reviewers concerns. Page 4: I would argue against the label ‘internal validity’, in the usual prediction modelling sense, evaluating a model on the same data used to derive the model is apparent validity, whilst conducting a cross-validation (or bootstrapping) to evaluate the model is internal validation and in fact table 1 reports the model and the associated model fit diagnostics, which I’m sure I would refer this as internal model fit, just ‘model fit diagnostic’, a minor point, but nomenclature in prediction modelling is tricky and thus labelling this ‘internal’ I fear would be confusing. Page 7 (lines 22-23) I don’t think I understand ‘The count of cases was entered as a variable with missing values, which we could predict, given the model coefficients and population at risk’ - can the authors elaborate (what do you mean ‘with missing values’?) Page 24: Can the authors explain ‘Number correct’

VERSION 2 – AUTHOR RESPONSE

Reviewer 1 (John McGrath)

Thank you for your positive comments in light of our initial changes to the manuscript

Reviewer 2 (Thomas P.A. Debray)

Thank you for taking the time to review our manuscript.

1. We have clarified the number of strata included in the dataset (N=2,536) (see first paragraph page 7). We believe this is sufficient to have conducted k=10 cross-validation. We have chosen not to present a summary of these strata in an appendix because these strata are already summarised in the manuscript (for example, in the methods we detail the number of age, sex, ethnic groups and local authority regions in the prediction data which give rise to the number of unique strata). We also provide detail about the population at-risk, number of cases of psychosis from that population and data about the population density and other environmental characteristics in tables throughout the manuscript). Several previous publications from our group using these data have explored incidence rates across these strata in more detail, and we have cited these accordingly.

2. We thank the reviewer for this useful point. Understanding the external validity of our predictions to specific strata in the external validation sample is important. We have addressed this in the revised manuscript, by additionally inspecting how well our best-fitting prediction model (Model 7) performed for other subgroups, including by sex, age and ethnicity. (See the new Supplemental Table 2, the new paragraph in the results section (middle of p12) and the comment in the discussion (p14). This important point has helped us highlight where our model performs well, and where we need to concentrate future efforts to improve the validity of our model. We are grateful to the reviewer for this point. We took this opportunity to update the observed SEPEA cases included in the manuscript: since the study is ongoing, a small number of cases (n=6) have had a change in case status since the time we first submitted this manuscript. Five subjects have been excluded as they did not meet criteria for FEP, while 1 case became eligible. We have updated Tables 2-3 to reflect these changes, which do not alter the overall interpretation of the prediction models.

3. See our comment to point 2. We have attempted to clarify the external validity of our model in specific strata.

4. We have made this adjustment to Table 1 (see page 23).

Reviewer 3 (Gary Collins)

We are grateful to the reviewer for his thoughtful comments.

Page 4: The reviewer is correct regarding the mistake in nomenclature in our manuscript. We have altered “internal” validity to “apparent” validity throughout the paper. We have retained the use of the word “internal” in other contexts to differentiate those analyses from analyses applied to i.e. extrapolated to the external population of our validation area i.e. the SEPEA study

Page 7: We have attempted to clarify this. In order to predict the count of cases in the population at risk of our new region we need to specify a vector in the dataset where the predicted count of cases will be stored when the risk coefficients from the model are applied to the population at-risk. To do this, we set up a vector (i.e. a variable) with missing data. This field is populated with the count of cases in each stratum when we run the model. From this data we can sum the predicted count of cases by region, sex, age etc… We have clarified this point in the text (see Page 7, second

paragraph).

Page 24: We have clarified what we meant by “Number correct” in Table 3 (see P25). We have changed this to “Observed case count within SEPEA overall prediction intervals?” This reports whether the total observed number of cases falls within the prediction intervals given by the model.

General change to manuscript

We made a small revision to the “Article Summary” box to include an important point in the “key messages” section on interpreting our models for EIS planning. We deleted a previous point in this box to make room for this new point, and made a further small edit to the “strengths and limitations” section of the article summary to reflect the “deleted” point.