Predicting the development of schizophrenia

Chuma & Mahadun\(^1\) report on a much needed and topical meta-analysis of prospective studies investigating the predictive validity of prodromal criteria in schizophrenia. The potential importance of early identification and treatment cannot be underestimated. The authors should be congratulated for helping clarify whether the identification component is currently worthwhile. I have no doubt that this paper is generally well conducted and for the ‘ultra-high-risk strategy’ sample size reasonable but I am afraid I cannot agree with their interpretation of results. In particular, they conclude that both ultra-high-risk and basic-symptoms criteria are valid and useful tools in predicting the future development of schizophrenia among the ‘at-risk population’, and that ultra-high-risk criteria were able to ‘correctly predict schizophrenia’ (citing sensitivity of 81%), while being able to ‘exclude this condition with some certainty’ (citing specificity of 67%). Taken at face value, clinicians would conclude that these methods both rule in those who are going to develop schizophrenia and rule out those who will not develop schizophrenia with high certainty. A small point, but sensitivity relates more closely the ability to rule out a condition (and is linked with negative predictive value) and specificity to ruling in a condition; hence the Sacket acronym SP-in and SN-out. In black and white terms, a specificity of 67% immediately suggests there will be a problem with false positives. But neither sensitivity nor specificity is a substitute for positive predictive value and negative predictive value which are the actual accuracy rates for every person identified as at high risk (screen positive) or low risk (screen negative) by these tools after taking into account the conversion rate. I am uncertain why the authors have presented clinically obscure statistics like DOR but omit the informative ones, namely positive predictive value (PPV)/negative predictive value (NPV).

Using the pooled estimate of 81% sensitivity and 67% specificity and a conversion rate of 21% (402 of 1918 at baseline), the PPV of the ultra-high-risk method(s) would be 39.4%, meaning only four out of ten identified as ‘will progress to schizophrenia’ actually would do so, and six would not. Of course we do not know whether others would progress if we extend the follow-up period but this is currently speculation requiring re-examination of these tools over a longer period. Hypothetically, if 30% of people progressed, then the PPV of ultra-high-risk method(s) could rise to 50%, which is still disappointing in my opinion. More encouragingly perhaps, even at 21%, the NPV would be 93.0%; meaning almost 19 out of 20 thought to be at low risk would not progress. The numbers for those using basic symptom criteria are similar but with even better NPV (PPV = 38.6%; NPV = 98.7%). That said, it is not initially obvious that only 60–70% of people who will not convert are put in a low-risk category by the tool (i.e. it is redundant in a third) and basic symptom data come from only one study with 160 participants. I appreciate that many might find these statistical terms confusing. Previously, I have proposed a simple adjustment of false positives and false negatives per every 100 patients seen, which I called real-world interpretation/yield. So, for every 100 individuals thought to be at risk and subject to ultra-high-risk criteria, 17 would be correctly classified as converters to schizophrenia and 4 would be missed; and 53 would be correctly classified as non-converters but 26 would be falsely identified. In effect, there would be six times as many false positives as false negatives. If each ‘positive’ were treated, then (by ratio of false positives to true positives) 50% more patients without any prospect of psychosis would be treated than those actually at risk of psychosis. I wonder whether these error rates are really acceptable when mental health resources are stretched and long-term adverse effects of antipsychotics are more than ever before seen as problematic. I therefore ask the authors to reconsider whether these approaches are entirely valid for both rule-in and rule-out purposes when the data suggest mainly the latter. I also suggest a novel future study in which clinicians working with high-risk patients are randomised to predicting risk with and without the tools, a method that would elucidate the ‘added value’ in clinical practice.


Correspondence

Edited by Kiriakos Xenitidis and Colin Campbell

Predicting the development of schizophrenia

Chuma & Mahadun\(^1\) report on a much needed and topical meta-analysis of prospective studies investigating the predictive validity of prodromal criteria in schizophrenia. The potential importance of early identification and treatment cannot be underestimated. The authors should be congratulated for helping clarify whether the identification component is currently worthwhile. I have no doubt that this paper is generally well conducted and for the ‘ultra-high-risk strategy’ sample size reasonable but I am afraid I cannot agree with their interpretation of results. In particular, they conclude that both ultra-high-risk and basic-symptoms criteria are valid and useful tools in predicting the future development of schizophrenia among the ‘at-risk population’, and that ultra-high-risk criteria were able to ‘correctly predict schizophrenia’ (citing sensitivity of 81%), while being able to ‘exclude this condition with some certainty’ (citing specificity of 67%). Taken at face value, clinicians would conclude that these methods both rule in those who are going to develop schizophrenia and rule out those who will not develop schizophrenia with high certainty. A small point, but sensitivity relates more closely the ability to rule out a condition (and is linked with negative predictive value) and specificity to ruling in a condition; hence the Sacket acronym SP-in and SN-out. In black and white terms, a specificity of 67% immediately suggests there will be a problem with false positives. But neither sensitivity nor specificity is a substitute for positive predictive value and negative predictive value which are the actual accuracy rates for every person identified as at high risk (screen positive) or low risk (screen negative) by these tools after taking into account the conversion rate. I am uncertain why the authors have presented clinically obscure statistics like DOR but omit the informative ones, namely positive predictive value (PPV)/negative predictive value (NPV).

Using the pooled estimate of 81% sensitivity and 67% specificity and a conversion rate of 21% (402 of 1918 at baseline), the PPV of the ultra-high-risk method(s) would be 39.4%, meaning only four out of ten identified as ‘will progress to schizophrenia’ actually would do so, and six would not. Of course we do not know whether others would progress if we extend the follow-up period but this is currently speculation requiring re-examination of these tools over a longer period. Hypothetically, if 30% of people progressed, then the PPV of ultra-high-risk method(s) could rise to 50%, which is still disappointing in my opinion. More encouragingly perhaps, even at 21%, the NPV would be 93.0%; meaning almost 19 out of 20 thought to be at low risk would not progress. The numbers for those using basic symptom criteria are similar but with even better NPV (PPV = 38.6%; NPV = 98.7%). That said, it is not initially obvious that only 60–70% of people who will not convert are put in a low-risk category by the tool (i.e. it is redundant in a third) and basic symptom data come from only one study with 160 participants. I appreciate that many might find these statistical terms confusing. Previously, I have proposed a simple adjustment of false positives and false negatives per every 100 patients seen, which I called real-world interpretation/yield. So, for every 100 individuals thought to be at risk and subject to ultra-high-risk criteria, 17 would be correctly classified as converters to schizophrenia and 4 would be missed; and 53 would be correctly classified as non-converters but 26 would be falsely identified. In effect, there would be six times as many false positives as false negatives. If each ‘positive’ were treated, then (by ratio of false positives to true positives) 50% more patients without any prospect of psychosis would be treated than those actually at risk of psychosis. I wonder whether these error rates are really acceptable when mental health resources are stretched and long-term adverse effects of antipsychotics are more than ever before seen as problematic. I therefore ask the authors to reconsider whether these approaches are entirely valid for both rule-in and rule-out purposes when the data suggest mainly the latter. I also suggest a novel future study in which clinicians working with high-risk patients are randomised to predicting risk with and without the tools, a method that would elucidate the ‘added value’ in clinical practice.

The aim of Chuma & Mahadun’s study\(^1\) is compelling and of great clinical interest for preventive interventions in psychosis. Unluckily, the results presented by the authors are not reliable as they are undermined by severe methodological caveats.

First, the systematic research of the literature failed to uncover the majority of studies reporting follow-up transition rates in a sample of individuals at high risk (HR) for psychosis. The authors included only 12 studies but many more were available in the electronic databases (references supplied). Second, the authors did not check for potential overlapping between samples including studies enrolling the same individuals. For example, Yung et al\(^2\) and Yung et al\(^3\) were both retrieved despite the authors of these studies clearly stating in their manuscript that the ‘current paper’\(^4\) continues that research \([n = 49]\) by expanding the sample size to 104. The same applies to Yung et al\(^5\) and Yung et al\(^6\) while Woods et al\(^7\) is a revised analysis of Cannon et al\(^8\). Third, the authors stated in the inclusion criteria that the included studies ‘had a clearly specified population, from which a prodromal criterion was administered to identify clearly those with prodromal symptoms [HR+] from those without [HR –]’. There is no such a ‘clearly specified population’ from which the high-risk individuals are sampled. The sampling is based on help-seeking behaviours and does not epidemiologically represent the local population; in fact, the prevalence of high-risk symptoms in the general population is unknown. The second requirement to be included in the Chuma & Mahadun’s meta-analysis was that ‘the two groups [HR+ and HR –] were then followed up for a number of months and assessed again with a diagnostic instrument to determine those who had converted to schizophrenia’. This is really surprising as the vast majority of the longitudinal studies
did not follow-up the help-seeking individuals who underwent the clinical assessment at the prodromal services but were not considered at risk for psychosis (HR–). Consequently, it is completely obscure how the authors may have estimated the correct prevalence of false negatives (HR–, who developed psychosis over time) in their analysis. Given all the above concerns, I feel the results of this meta-analysis should be considered carefully as pilot data strongly undermined by significant methodological biases.


We also agree that assessing the clinical usefulness of prodromal criteria requires further consideration. We plan to further examine this important question in a subsequent paper. We welcome Dr Mitchell’s proposal for a randomised study where high-risk patients are randomised to predicting psychosis with or without formal tests for prodromal criteria.


Abortion, mental health and charges of guilt by association

Coleman’s meta-analysis of abortion and mental health studies was harshly criticised in three letters by five authors (Robinson, Stotland, Nadelson, Coyne, and Littell) who all cited an Ethics & Medicine article I wrote (not Coleman) as evidence that Coleman’s study cannot be trusted. My full response is summarised as follows.

First, Robinson’s assertion that I am Coleman’s ‘leader’ is nonsense. We have no institutional, financial or personal entanglements. Second, I gathered data that required the analysis of research psychologists. I am thankful that Coleman agreed to analyse it and help present it in a scientifically accurate and impartial manner. As a biomedical ethicist, I explore the intersections of medicine, science, philosophy, theology, ethics and the law. When writing papers intended for each of these fields, I seek to use the language and tools appropriate to each field.

Third, the cited article was a response to a pro-life philosopher who argued that any evidence of emotional suffering of women following abortion is essentially irrelevant to the moral argument against abortion and counterproductive to pro-life efforts. The core of my response was that Christians have an obligation to seek to use the language and tools appropriate to each field.

Fourth, it also reflected my sincere belief that abortion involves substantial dangers to specific subgroups of women. Unfortunately, critics have distorted this into the charge that I seek to scare women with exaggerated risks. That is untrue. There are real risks, especially for certain higher-risk groups. Women
Correspondence

should be told of the truth regarding statistically significant findings. These should be neither exaggerated nor minimised.

Finally, women who dare to express emotional trials following an abortion face rejection from people on both sides. A few pro-lifers harshly dismiss these women as ‘sinner’s who deserve a lifetime of grief. Conversely, at least a few pro-choicers dismiss their grief as ‘whining’ or ‘rare’, or suggest that only women mentally unstable prior to their abortions would complain so much. By contrast, the post-abortion healing movement simply asks those on both sides to respect the experiences of women grieving a past abortion. But even this pro-healing position is attacked. Pro-choicers accuse us of manipulating gullible women into falsely blaming unrelated life problems on their abortions.\(^6\) Some pro-life advocates, meanwhile, accuse us of encouraging an unprincipled, narcissistic worldview that diminishes the moral absolutes regarding the sanctity of life.\(^5\)

To my mind, the question of whether abortion is the sole, direct cause of certain mental illnesses is far less important than the fact that many self-aware women want help coping with a past abortion experience.\(^7\) Why is it so hard to simply accept their self-assessments and stated needs? Women deserve better.

Declaration of interest

D. R. is Director of the Elliot Institute, USA.


doi: 10.1192/bjp.200.3.255a

The fall of the atypicals?

In his editorial arguing that atypical antipsychotics can no longer be regarded as having advantages over conventional drugs, Kendall\(^1\) makes two statements which do not do justice to the available evidence.

First, he states: ‘With the exception of clozapine for treatment-resistant schizophrenia, the atypicals, as a group of antipsychotics, are no more efficacious for schizophrenia than the typicals, whether it is chronic or acute, for first or subsequent episodes, for the acute episode or for promoting recovery’. This is supported by a reference to the updated National Institute for Health and Clinical Excellence (NICE) guideline for schizophrenia,\(^2\) which in turn bases its conclusions on a series of meta-analyses carried out by the National Collaborating Centre for Mental Health (NCCMH; www.nccmh.org.uk). The problem here is that two other meta-analyses have reached different conclusions. In 2003, Davis et al\(^3\) found that, apart from clozapine, three atypicals showed significant superiority over conventional antipsychotics: risperidone (22 studies, effect size (ES) 0.25), olanzapine (14 studies, ES = 0.21) and amisulpride (12 studies, ES = 0.29). Six years later, Leucht et al\(^4\) had closely similar findings for olanzapine (28 studies, ES = 0.28) and amisulpride (13 studies, ES = 0.31); the effect size had become smaller for risperidone, but it was still significant (34 studies, ES = 0.13).

One reason why the NICE/NCCMH meta-analysis may have reached negative conclusions concerning these three drugs is that it included fewer studies. The outcome measure used by Davis et al\(^3\) and Leucht et al\(^4\) was reduction in total symptom scores, based on pooled data from the Positive and Negative Syndrome Scale (PANSS), Brief Psychiatric Rating Scale (BPRS) and CGI (Clinical Global Impression) scale, and from the PANSS and BPRS respectively. In the NICE/NCCMH meta-analyses there were only 17 studies of risperidone, 10 of olanzapine and 4 of amisulpride in which the drug was compared with a conventional antipsychotic on any of these outcome measures. Data were also pooled separately for studies carried out on patients with first-episode schizophrenia, on those with acute exacerbations or recurrences, and on those with treatment-resistant illness. As a result, the maximum number of studies included in any of the NICE/NCCMH meta-analyses of overall symptoms for these three atypicals was six, and several contained only one or two studies.

Later in the article, Kendall cites approvingly a meta-analysis by Geddes et al\(^5\) in 2000, which found evidence that the superiority of atypicals (including clozapine) was an artefact of the high dose of the typical antipsychotic used as a comparator in some of the trials. These authors used meta-regression to examine the predictive value of haloperidol dose (23 studies) or chlorpromazine dose (7 studies) on the outcome of total symptom scores. In both cases, the findings were significant: an observed advantage in favour of atypicals disappeared as the dose of the comparator drug decreased. Davis et al\(^3\) subsequently explored the effect of comparator dose in their meta-analysis. The results of several different analyses led them to conclude that there was no significant effect of haloperidol in a larger data-set of studies. Leucht et al\(^4\) also failed to find a significant effect of chlorpromazine comparator dose in another meta-analysis carried out at around the same time. Geddes et al\(^5\) argued that a significant effect of comparator dose could be re-instated in this latter meta-analysis by using their own meta-regression technique; in their author reply, Leucht et al\(^4\) countered that the effect was not significant when a variety of other statistical techniques were used, indicating that the finding was not robust.

Kendall states that the comparator drug effect has been ‘neither confirmed nor disproved by later meta-analyses’. An arguably more accurate conclusion is that it was an early finding which has not stood the test of time.


3 Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. Arch Gen Psychiatry 2003; 60: 553–64.


**Author’s reply:** Dr McKenna seems to have misread and misunderstood the editorial.1 I do not argue that ‘atypical antipsychotics’ (whatever they are) can no longer be regarded as having advantages over ‘conventional drugs’ (whatever they are). I argue that the class – the ‘atypical’ antipsychotics – has been fabricated for marketing purposes and has no basis in science or clinical practice. Although some studies do suggest that individual drugs differ in terms of side-effects, potency, efficacy and effectiveness, the differences – with the exception of clozapine for treatment-resistant schizophrenia – are small, and their relative effects are, at least in part, dependent on the potency2 and dose3 of the comparator. These differences do not constitute a ‘class effect’.

In the meta-analyses for the schizophrenia NICE guideline,4 we examined the use of antipsychotics in a number of different clinical contexts (e.g. first episode, acute episode and treatment resistance) and concluded that the differences in efficacy between drugs were unlikely to be clinically important. However, the guideline did acknowledge, as do other meta-analyses,5,6 that differences in terms of side-effects allow clinicians and service users to find a drug that suits them. Moreover, all three meta-analyses agree that there are no consistent differences or similarities between ‘typicals’ and ‘atypicals’— this is an important perspective that McKenna seems to have missed.

In undertaking our meta-analyses for the development of a guideline, we were guided by a broad range of clinical review questions. The more specific the question the fewer studies are likely to be able to answer the question. The data underpinning the use of antipsychotics in the treatment of acute schizophrenia included over 72,000 patients, whereas for the first episode this figure dropped below 2000. We could have lumped more data together: it is very unlikely that increasing the numbers of studies and participants with different presentations in the meta-analyses would change the central conclusions (that oral antipsychotics are all much the same in terms of efficacy); but it would have significantly diminished the clinical utility of each analysis.

The study by Geddes et al3 is important not only in highlighting the influence of the comparator dose on efficacy, but also in questioning the integrity and claimed superiority of the class of ‘atypicals’. It is true that Davis et al5 did not confirm the findings of Geddes et al3; nevertheless, I maintain that the findings have clinical face validity. Not irrelevant to this perspective is that Leucht et al2 in their paper summarising the debate, said ‘It is a major limitation that only a few studies used mid-potency FGA [first-generation antipsychotic] comparators. We recommend that each new drug is compared with a low-potency, a mid-potency, and a high-potency FGA.’ Explicit in this recommendation is that the potency of the comparator can introduce bias; it would be odd to suggest that the dose of the comparator would not also have an important effect. In any event, McKenna may be in danger of not seeing the wood for the trees: the ‘atypicals’ have surely fallen.

---

**Correspondence**

5 Davis JM, Chen N, Glick ID. A meta-analysis of the efficacy of second-generation antipsychotics. Arch Gen Psychiatry 2003; 60: 553–64.
Predicting the development of schizophrenia
Paolo Fusar-Poli
Access the most recent version at DOI: 10.1192/bjp.200.3.254a

References
This article cites 7 articles, 2 of which you can access for free at:
http://bjp.rcpsych.org/content/200/3/254.2#BIBL

Reprints/permissions
To obtain reprints or permission to reproduce material from this paper, please write to permissions@rcpsych.ac.uk

You can respond to this article at
/letters/submit/bjprcpsych;200/3/254-a

Downloaded from
http://bjp.rcpsych.org/ on June 15, 2017
Published by The Royal College of Psychiatrists