Author’s response to reviews

Title: Prospective study of factors influencing conditional discharge from a forensic hospital: the DUNDRUM-3 programme completion and DUNDRUM-4 recovery structured professional judgement instruments and risk.

Authors:

Mary Davoren (davorenm@tcd.ie)
Zareena Abidin (zareena.abidin@hse.ie)
Leena Naughton (leenanaughton@gmail.com)
Olivia Gibbons (olivia.gibbons@yahoo.com)
Andrea Nulty (andrea.nulty@hse.ie)
Brenda Wright (brenda.wright1@hse.ie)
Harry G Kennedy (kennedh@tcd.ie)

Version: 3 Date: 18 April 2013

Author’s response to reviews: see over
Thank you. We are grateful for the comments of all three reviewers and we hope that we have addressed the reviewer’s comments as follows -

Reviewer 1:
Sandy Simpson
Reviewer’s report:
The paper is much improved and I commend the authors for their responses to the prior reviews. The longer time period of follow up has also benefited the study. My comments are now minor.
Major compulsory revisions: nil

Minor essential revisions: The study is of 56 eligible patients. Figure 1 is unnecessary, as is much of the first para, page 8.

We have now omitted figure 1 and edited paragraph 1 page 8 as follows -

“All patients in the hospital patients were assessed between February and March 2011 of whom 56 were newly eligible for conditional discharge to the community by the mental health review board. Data on the eligible patients and their subsequent discharges from the hospital were then gathered up to 31st December 2012 (Fig. 1). Those detained because UST (8 patients) were eligible for conditional discharge by the MHRB in the same way as those detained because found NGRI (48 patients), though those detained because UST could also be returned to court by the treating consultant if they regained fitness. “

Discretionary revisions: If I have read the paper accurately, this is a study of 56 patients each experiencing 3 reviews by the MHRB, for a total of 168 decision points.

Correct

The authors have used measures [the DUNDRUM and other tools] taken in Feb-March 2011 to predict release by December 2012, yet some of the tools are dynamic and may well have changed between the first and third MHRB hearings, for instance. Could the authors comment on the amount of change that the DUNDRUM 3 and 4, and indeed the dynamic components of the other measures might have taken place of over this time period, and whether that has any influence on the predictive power of measures possibly taken 18 months prior to the final release opportunity?

The reviewer has correctly read and understood the study design. Because we wanted a prospective design, we rated all patients prior to the introduction of new legislation, with ratings carried out by a researcher. The clinicians and the MHRB were blind to the research ratings presented here. Many of the scales and items rated are dynamic and would have changed over time. This would tend to attenuate the predictive power of the ratings – we have now included this point in the ‘limitations’ section of the discussion. In a forthcoming study, we will examine whether dynamic change in the various
rating scales reflects or predicts change in outcomes. We believe this is best dealt with in a different study design.

“A further limitation of this study is that many of the measures are dynamic. Ratings made up to 18 months before a hearing may therefore be attenuated by subsequent change. Significant associations (receiver operating characteristics, odds ratios) are therefore conservative estimates”.

Level of interest:
An article of outstanding merit and interest in its field
Quality of written English:
Acceptable
Statistical review:
Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I declare that I have no competing interests.

Reviewer 3:
Philip Sugarman
Reviewer's report:
none
Level of interest:
An article of importance in its field
Quality of written English:
Acceptable
Statistical review:
No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
'I declare that I have no competing interests'

Reviewer 3:
Michael J Vitacco
Reviewer's report:
The paper is generally well written and organized. The primary issue with the paper is addressed in the compulsory section.
I was pleased to review the paper entitled, “Prospective study of factors influencing conditional discharge from a forensic hospital: The DUNDREUM-3 programme completion and DUNDREUM-4 recovery and structured professional judgment instruments and risk.” This is a re-review, and I was privileged to read the initial submission of this paper. Overall, the authors were receptive to the reviews and their work is evidenced in a much improved paper. There are still some issues that need to be addressed:
1. In describing diagnoses, the authors appear to use several that are not official diagnoses. For instance, psychotic depression would appear to be Major Depression with Psychotic Features.
Thank you for this reminder. We have now listed diagnosis by ICD-10
categories. We have had to include the term 'mental retardation' as per ICD-
10 although this is a term objected to by service users in Ireland and
elsewhere on the grounds of stigma. We have therefore also used the legal
category ‘intellectual disability’ which is more acceptable to service users and
carers.

“Diagnosis according to ICD-10 criteria [22] was schizophrenia (F20) 41 (73%),
schizoaffective disorder (F25) 4 (7%), bi-polar affective disorder (F31) 5 (9%), recurrent
depressive disorder, severe with psychotic symptoms (F33.3) 4 (7%), intellectual disability
(mental retardation with significant impairment of behaviour F71.1 and 72.1) 2 (4%). The
legal status was unfit to stand trial 8 (14%), not guilty by reason of insanity 48 (86%)”. P12

2. My primary concern remains with the outcomes and potential confounds with
the independent variables. The fact that these outcomes are not independent is a
primary concern. More to the point, it is my opinion the authors still have not
properly addressed this confound. And, this is underscored by logistic
regressions predicting 95.1% of the outcomes. This pattern is repeated
throughout the logistic regressions. The fact the variables are so intertwined
limits the study to a significant degree.

We regret that reviewer 3 appears to have misread this paragraph and the following paragraph. The
95.1% figure cited by reviewer 3 refers to those correctly identified as “not discharged” in the first
model. The passage (see below) clearly states that only 83.1% of those who were discharged were
correctly classified, 92.5% correctly classified overall (see below, emphasis added). The first
regression analysis model referred to by the reviewer (“95.1%” etc) demonstrated that the GAF was
the only significant measure in the final model. The GAF rating was not available to the clinicians
reporting to the MHRB or to the MHRB itself and could not have directly influenced the outcome.
The GAF is however an ‘over view’ variable which is obviously related to many facts that in turn
influence decision making. This is the purpose of the study and we have discussed this in the final
section. Furthermore, the model the reviewer refers to was not robust and backward regression
yielded a different result, a point that we carefully explained in the following passage pp15-16 (see
below). The final model presented correctly predicted only 66% of those discharged and 88% of
those not discharged, an overall correct prediction of 95%. The most important statistic here is the
correct prediction of discharges (66%) not the overall 95%. The relatively low prediction of those
discharged shows that this is not an over-determined or over-confounded model. Please note that to
meet this reviewer’s further comments, we have taken the word ‘predicted’ out of this and all other
passages, although measures made at baseline that statistically relate to an outcome could be
regarded as ‘predictions’ where the outcome occurred blind to the measure.

The iterative process resolved in one step for the GAF as the only significant variable
remaining in the regression equation. The model was acceptable with Hosmer and Lemeshow
$X^2=8.45, df=8, p=0.391$, Nagelkerke $R^2 = 0.610$, and GAF odds ratio=1.258 (95%
confidence interval 1.089 – 1.454) $p=0.002$. This model correctly distinguished 95.1% of
those not discharged, 83.1% of those who were discharged and 92.5% overall. This model
was not robust with backwards logistic regression yielding a different result, resolving in five
steps (DUNDRUM-3 odds ratio=0.682, 95% CI 0.470 – 0.991, p=0.045; SAPROF OR = 0.633, 95% CI 0.413 – 0.967, p=0.034; GAF OR = 1.387, 95% CI 1.054 – 1.825, p=0.020, Hosmer and Lemeshow $X^2 = 3.629, df=8, p=0.889$, Nagelkerke $R^2 = 0.739$, correct predictions as before).

Because the model was not robust, the logistic regression was repeated, this time omitting the GAF. This was done because of the apparent interactive effects between the GAF and SAPROF, evident also when the START-S was included instead of the SAPROF. Backward logistic regression resolved in five iterations, Nagelkerke $R^2 = 0.603$, Hosmer and Lemeshow $X^2 = 6.05, df=8, p=0.642$, with 97.6% correct for those not discharged, 66.7% correct for those discharged and 90.6% correct overall. The model included two variables DUNDRUM-3 OR = 0.717, 95% CI 0.561 – 0.916, p=0.008 and PANSS total score which did not reach significance OR = 0.913, 95% CI 0.821 – 1.016, p=0.097. This model was reasonably robust and when repeated by forward logistic regression yielded a model in one iteration with Nagelkerke $R^2 = 0.525$, Hosmer and Lemeshow $X^2 = 2.37, df=7, p=0.927$, 92% correct for those not discharged, 66.7% correct for those discharged and 86.8% correct overall and included a single variable, DUNDRUM-3 odds ratio = 0.698, 95% CI 0.563 – 0.866, $p<0.001$. pp14-15

This is not primarily an actuarial study. In this study, we have no clinical interest in predictive scores apart from their use as one part of validation. The purpose is to validate the content of structured professional judgement instruments. As the other reviewers recognised, the MHRBs were blind to the instrumental ratings made prior to the commencement of the new legislation. They were not blind to the clinical facts of the individual patient’s progress. The object of the study is to examine whether the content of the instruments captures the clinical information that influences the decision makers. We have now made this more explicit at several points in the text.

“This 'experiment of nature' offered the opportunity to carry out a prospective observational study on the clinical facts that influence how clinicians make such recommendations and MHRBs make such decisions. “p5

“The clinicians and MHRB were blind to the assessments reported here, carried out independently of the treating clinicians prior to the commencement of the new legislation. However neither the clinicians nor the members of the MHRB could be blinded to the factual unstructured clinical information used to make the ratings for the various instruments, nor could they have been blinded to such information. Clinicians decided to recommend or not to recommend absolute or conditional discharge in accordance with their normal practice, a synthesis of the results of structured professional judgement instruments such as the HCR-20 [rated by themselves] and broad bio-psycho-social assessment including reports of progress in hospital. The MHRB received reports from the treating psychiatrists and other clinicians and was independent in the exercise of its powers [15]”. Pp6-7

“….These items are often included in clinician’s unstructured reports to mental health tribunals and review boards to assist these bodies in their decision making with regard to a patient’s readiness for conditional or absolute discharge, or neither.” P9 (as before)
“The treating clinicians were blind to these audit and outcome ratings when making their reports to the MHRB, as were the MHRB.” p10

“Association with outcome was tested using the receiver operating characteristic (ROC) area under the curve (AUC),” p11

“Regression analysis was used to examine the antecedent covariates of discharge.” p11

“This was not an actuarial study. This study was not primarily about predicting discharge though this forms one part of validation. The main purpose was to validate the content of structured professional judgement instruments to aid decision making regarding discharge from a forensic hospital. The MHRBs were blind to the instrumental ratings made prior to the commencement of the new legislation. They were not blind to the unstructured clinical facts of the individual patient’s progress. The object of the study was to examine whether the content of the instruments captured the clinical information that influenced the decision makers. A successful validation would be evidence from receiver operating characteristics and in regression models that the instruments accounted for most of the outcomes. This proved to be the case. We believe the study demonstrated that simple risk assessments account for only some of the outcome, with treatment completion and recovery measures also contributing. Further, we wanted to examine whether all items in these structured professional judgement instruments were relevant. We found that some instruments that had scale scores associated with conditional discharge had only a few items that were relevant to this outcome, though other items might be relevant to other outcomes such as later reoffending or recall to hospital. Odds ratios also demonstrated that not all items were equally influential, some having larger effects (stronger associations with outcomes) than others.” p17

“It is particularly interesting that the Global Assessment of Function (GAF) emerged as one of the strongest precursors of conditional discharge, and dominated the logistic regression models. The DUNDRUM-3 and DUNDRUM-4 correlated strongly with the GAF (Spearman r for DUNDRUM-3 and GAF, r = -0.804, p<0.001, DUNDRUM-4 and GAF r = -0.729, p<0.001) and the DUNDRUM-3 programme completion scale emerged as the next strongest statistical precursor. It is possible that the items of these scales represent elements of what is taken into account when rating the GAF, or it may be that an element in the rating of all of these items, like the GAF, is a rating of some ‘structured intuition’ as recently suggested [28]. The ratings of the seven items of the DUNDRUM-3 programme completion scale and the six items of the DUNDRUM-4 recovery scale have in common that they are based on motivation theory, cycle of change and engagement.” p19

3. The first paragraph in the limitations section is highly confusing. I understand you are talking about different populations but the fact is the MHRBs were not blinded to the instruments, and that means this study is not predictive. Based on this limitation, it is my opinion these results should not be presented.
The reviewer has perhaps misunderstood the design and purpose of the study as we have presented it. This is not an actuarial study. We have no practical interest in predictive scores, though this is one part of validation. The purpose is to validate the content of structured professional judgement instruments. As the other reviewers recognised, the MHRBs were blind to the instrumental ratings made prior to the commencement of the new legislation. They were not blind to the clinical facts of the individual patient’s progress. The object of the study is to examine whether the content of the instruments captures the clinical information that influences the decision makers. We have now made this more explicit at several points in the text.

We have redrafted the first paragraph of the ‘limitations’ section. We hope this is now clearer.

The ratings of the instruments and their component items were carried out by researchers who were blind to each others’ ratings and the research ratings were not provided to the clinicians or MHRBs. The clinicians and MHRB were therefore blind to the baseline ratings. The success or failure of the instruments as statistical predictors depends on the extent to which they emulate the ways in which the clinicians and MHRBs make their recommendations and decisions. The clinical information contained in the structured professional judgement instruments was inevitably available in unstructured form to the clinicians and MHRBs who could not be deprived of it. What was studied is the extent to which the instruments and their contents at baseline, were associated with the later behaviour of the decision makers. The statistical inference is that the decision makers relied to a greater or lesser extent on this information. This was an observational study and therefore no special effort could be made to blind clinicians who gave evidence or the members of the MHRB to such unstructured information. Since the purpose of the study was to identify in a defined structured form those clinical facts that influence conditional discharge, it would be self-defeating and probably impossible to blind clinicians and MHRB members to such unstructured information. P20