

The Incidence of Treatment Resistant Schizophrenia using the TRRIP Definitions (3rd Version)

Dan W. Joyce, Alexis E. Cullen, Rodrigo A. Bressan and Sukhwinder S. Shergill

April 15, 2018

Abstract

Importance

Estimates of treatment resistant schizophrenia (TRS) vary due to lack of consensus definition. The Treatment Response and Resistance in Psychosis (TRRIP) consensus provides a rigorous prospective definition for TRS, but has not yet been applied to data. We provide the first prospective estimate of the incidence of TRS in a large community cohort using TRRIP.

Objective

To establish the incidence of TRS prospectively in a “real world” community-dwelling population of people with chronic schizophrenia by data-mining with an operationalised algorithm implementing the TRRIP consensus.

Data Sources

The NIMH Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial dataset.

Main Outcome and Measure

Crude incidence rates and survival estimates for TRS, along with inferential analyses of baseline demographics, illness characteristics and medication predicting the probability of meeting the TRRIP criteria for TRS. Machine learning was also applied to these features in an attempt to identify robust predictors for TRS.

Results

From 1458 patients enrolled in CATIE, XX patients received two or more adequate treatments, and of these, XX developed TRS by the prospective TRRIP criteria, yielding a crude incidence of XX cases per 100 contributed person-years. There was no difference in medication, dose or sex between those who responded to treatment and those meeting TRS thresholds. Developing TRS was modestly associated with baseline positive symptom burden, age, and verbal memory; however, machine learning failed to validate these as robust predictors of developing TRS.

Conclusion and Relevance

The CATIE trial protocol excluded patients with retrospective evidence of TRS, however, applying the TRRIP definition revealed that there were patients meeting criteria *prospectively*. Our results suggest a XXX incidence, no effect of patient sex and that baseline clinical and demographic data is not a robust predictor of future resistance status.

1 Introduction

Currently, there is a paucity of empirical data on the prevalence and incidence of treatment resistant schizophrenia (TRS) with estimates yielded varying rates, ranging from around one-third (Lehman et al., 2004; Elkis, 2007) to 40-60% (Juarez-Reyes et al., 1995; Essock et al., 1996; Meltzer and Kostakoglu, 2001), with males expected to be more resistant to neuroleptic treatment (Szymanski

et al., 1995), and variation in results being attributable to differing diagnostic criteria, populations (inpatient versus community settings) and methodological differences e.g. retrospective versus prospective assessment (Dammak, 2013).

The Treatment Response and Resistance in Psychosis (TRRIP) Howes et al. (2017) consensus statement aims to provide a uniform definition and minimum criteria for TRS unifying 8 different national, professional body and clinical guideline definitions that all agree on the number of antipsychotic trials (i.e. at least 2 different agents) but differ on the definition of adequate trial, dose, adherence, illness severity and clinical definition of response.

A key component of the TRRIP consensus is that evaluation of treatment resistance be *prospective* and the important minimum criteria are (Howes et al., 2017):

- Current Symptoms: at least **moderate** severity on standardised instruments (e.g. PANSS) and duration of at least 12 weeks. Social and occupational functioning (SOF) should be measured using a validated scale and show at least **moderate** impairment.
- Adequate treatment: at least 2 treatments of **duration** at least 6 weeks at a **therapeutic dose** and with adherence of at least 80% doses with information furnished from at least 2 sources
- Symptom Domain: specify whether positive, negative or cognitive symptoms are treatment resistant
- Onset: specify early (within 1 year), medium-term (1-5 years) or late (greater than 5 years) after starting treatment

In principle, the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial protocol excluded patients with retrospective evidence for TRS “defined by the persistence of severe symptoms despite adequate trials of one of the proposed treatments or prior treatment with clozapine” (Lieberman et al., 2005). Patients in CATIE had been unwell for at least one year, and a majority

(72%) were on treatment at the start of the trial (Lieberman et al., 2005). Additionally, CATIE is a large (over 1400 patients), multi-site trial of community-based patients and estimates of incidence of TRS in this population can provide insight into disease burden in patients with chronic illness, but not necessarily under specialist inpatient care.

We aim to estimate the incidence of TRS in a large community dwelling population of patients with chronic schizophrenia, with the hypothesis that in the CATIE trial, applying a rigorous definition *prospectively* will expose cases of treatment resistance that developed during the trial (that may have been assumed to have been excluded).

REVISE Our secondary analyses estimated the proportion of patients who meet social and occupational functioning and/or symptoms criteria but who had trialled only *one* adequate treatment. Finally, we conducted exploratory analyses (using machine learning) to determine if prediction of TRS cases is possible given baseline clinical data in CATIE.

2 Methods

The publicly available CATIE data (Study ID: N01 MH090001-06, https://ndar.nih.gov/edit_collection.html?id=2081) were mined using bespoke implementations of algorithms that operationalise the minimum TRRIP consensus criteria. All data pre-processing and analysis were conducted in R version 3.4.1.

2.1 Missing Data – NEEDS REVISING

We retrieved a total of 1458 patients from the CATIE data set. Complete baseline demographic data was available for 1380 patients. Of these 1380 patients, there were no missing psychopathology (PANSS) data, but 9 patients were missing adequate information to assess SOF. Therefore, 1371 patients could be included in our analysis. Of 1371 patients assessed at baseline, 1271 had sufficient

data for follow-up (100 patients were missing details of either medication trials, symptom or SOF data). To establish if follow-up data was systematically and predictably missing in these 1371 patients, three separate logistic regression models were estimated for the probability of missing i) symptom, ii) SOF and iii) adequate medication trial data using the baseline data (including demographics) as predictors.

Social and Occupational Functioning Measures

In CATIE, a number of instruments were used in the assessment of social and occupational functioning (SOF). The TRRIP consensus recommends the use of a standardised tool e.g. the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) for which operational criteria are defined in the Personal and Social Performance (PSP) scale (Morosini et al., 2000). The CATIE trial used a number of different scales which probe SOF, so we aggregated a number of different items from these instruments to arrive at a threshold for “moderate impairment” (see Supplementary Information for details).

Medication Doses, Duration and Adherence Measures

For medication dosing, we noted that using a 600mg chlorpromazine equivalent – using established consensus methods including daily-dose equivalents (Leucht et al., 2016) – resulted in restrictively high adequate dose thresholds for many of the antipsychotics in the CATIE trial. The TRRIP recommendations also allow for adequate dosing defined as the target or midpoint of the dose range in each medication’s summary of product characteristics (SPC) (see Supplementary Information, Table S1 for thresholds used) and this more realistically reflects clinical prescribing behaviour. The TRRIP medication adherence criteria specify $\geq 80\%$ adherence based on pill counts, chart inspection and collateral information. As a proxy, we used the CATIE overall adherence score, which combines multiple sources of information (including pill counts and collateral) for which the highest level of adherence available is $\geq 75\%$. In Phase 3 of the

CATIE trial, participants were allocated to either one or a combination of two antipsychotics; in cases where two antipsychotics were given, either could count as an adequate trial (provided the medication qualifying as the adequate trial was different to that assigned in Phase 2). Duration of treatment was adequate if greater than 6 weeks.

To test for differences in prescribed doses, multivariate analysis of variance (MANOVA) was used as an omnibus test, with drug dose as multiple-dependent and group (treatment resistant and non-treatment resistant – TRS and NTR respectively) as independent variables.

Symptom Change and Response Thresholds – NEEDS RE-VISING

Before and after an adequate trial of medication, patients must continue meet both absolute symptom severity criteria as well as demonstrating less than 20% change in symptoms. Following TRRIP, the absolute symptom severity criteria for each patient where having two PANNS items at 'moderate', or one item at 'severe' in positive or negative domains. For symptom change at the end of a medication trial, we calculated the reduction from baseline (using the Leucht et al. (2009) formula) and applied the TRRIP criteria of treatment resistance being a reduction of less than 20% in both the overall PANSS score *and* the positive or negative domains individually. When a patient exited the trial, after at least 2 adequate trials, their last PANSS and SOF scores were used to evaluate absolute threshold and change in symptoms, and classified as either TRS (if still meeting criteria) or right-censored and considered non-treatment resistance (NTR).

The TRRIP consensus further suggests treatment resistance in the cognitive domain and neurocognitive testing was performed in CATIE, but only at 4 time points (including baseline and the end of the study). Five domains (processing speed, reasoning, working memory, verbal memory and vigilance) were measured in CATIE, and one composite (average) score recorded. In each domain, the

cognitive performance scores are normalised (i.e. *z*-scores) *within* the CATIE population, not relative to a control or normative population data (Keefe et al., 2006). Cognitive performance change can therefore be evaluated in the CATIE trial (between baseline and end of the study), but they cannot be used as a domain for treatment resistance (e.g. in the same way positive and negative symptoms are) because of paucity of data coincident with symptom evaluations and changes to medication.

Survival and Incidence Estimates – NEEDS REVISING

A patient was deemed at risk of TRS if they had ≥ 2 adequate trials of different antipsychotic medications such that in the CATIE trial design, patients in Phase 1B and those who transitioned directly from Phase 1 to 2 onwards can be prospectively evaluated. For at risk patients, we estimated a Cox regression with survival time defined as conversion to TRS or censoring (i.e. leaving the trial and being below threshold for TRS criteria) with sex as a predictor. A Mantel-Haenszel test was used to establish violation of the proportional hazards assumption. A χ^2 test for proportionality was used to examine if the male/female proportion in the at risk for TRS group differed from the CATIE population. We estimated the crude incidence rate as the sum of TRS events divided by the person-years contributed by all at risk patients up to either conversion or censoring. Survival analyses, hypothesis tests and plots were all produced using the R packages `survival` v 2.41-3 and `survminer` v0.4.0.

Burden of Individual TRS Criteria – NEEDS REVISING

To establish how prevalent the SOF, symptom and treatment criteria are in the CATIE population, patients in each phase of the trial (baseline, Phase 1/1A, 1B, 2 and 3) were evaluated for their meeting criteria *independently* (in contrast to meeting all criteria for TRS). Absolute numbers and proportions were calculated. We then defined a “difficult to treat” sub-population (distinct from the TRS group) of patients that met TRS criteria on symptoms and/or

SOF *but* whom received exactly one adequate treatment prior to exiting the trial. These patients represent those with unmet needs, i.e. symptomatically or in terms of impact on social and occupational functioning and who might warrant further treatment trials.

Predicting TRS from Baseline Data – NEEDS REVISING

We conducted an inferential analysis using logistic regression for establishing associations between TRS status and baseline features (37 variables including SOF, symptoms (as total PANSS positive, negative and general domains), cognitive testing (5 domains), medication on entering the trial, recent exacerbations/hospitalisations, age and duration of illness, presence of tardive dyskinesia, and race/ethnicity). We applied machine learning (ML) to test the *predictive* ability of the baseline data to predict TRS from the same baseline variables (described in Supplementary Information, Table S4), using penalised generalised linear models – glmnets (Friedman et al., 2010) – to simultaneously implement variable/feature selection and classification (TRS vs. NTR) within a repeated, stratified k -fold cross-validation routine with area under the receiver-operating characteristic curve (AUC) used to estimate predictive out-of-sample performance. The ratio of TRS:NTR cases is heavily biased toward NTR leading to established problems with imbalanced classification in ML (He and Garcia, 2009), so given the relatively small number of predictor variables (Blagus and Lusa, 2013) we applied the synthetic minority oversampling technique (SMOTE) to balance classes within each fold during repeated cross-validation (Chawla et al., 2002).

3 Results

3.1 Missing Data– NEEDS REVISING

The probability of patients’ symptom, SOF and adequate medication trials data being missing at follow-up was estimated using three logistic regression models

on the 1371 patients using baseline clinical and demographic variables (PANSS positive, negative, general symptoms; age; sex; race; ethnicity; accommodation and independent living status; marital status; years since first treated; total and recent hospital admissions; baseline medication prior to entering the trial and baseline at-risk status from the TRRIPP criteria). Holms-Bonferroni correction applied to each model. None of the predictors were significantly associated missing follow-up data for symptoms, SOF or medication trials.

3.2 Survival Analysis and Incidence – NEEDS REVISING

We were able to retrieve data for a total of 1458 (male:female = 1078:380) patients in the CATIE data set where – throughout their time in the trial – 145 (34 female) were at risk of TRS by virtue of having had ≥ 2 adequate trials of different antipsychotics.

Of the 145 at-risk patients, a total of 59.3% (86 of 145) converted to TRS. In the at-risk group, 86 (of 145) converted to TRS over a total of 144.79 contributed person-years, yielding a crude incidence of 59.40 cases per 100 person-years. These 86 patients represent just under 6% of the total trial population. For patients converting to TRS (ignoring sex), the median survival time was 1.06 years (95% CI = [0.94, 1.30]) A two-sample test for equality of proportions of male:female (111:34) in the TRS at-risk sub-population compared with the remaining general CATIE population (967:346) showed no significant difference ($\chi^2 = 0.43$, $df = 1$, p -value = 0.51, difference in proportions = 0.013, 95% confidence interval = [-0.022, 0.049]) likewise, in the at-risk patient group, the proportion of male:females (63:23) converting to TRS versus non-treatment resistance (NTR) was not significantly different ($\chi^2 = 0.87$, $df = 1$, p -value = 0.35, difference in proportions = -0.11, 95% confidence interval = [-0.310, 0.093]).

Cox regression showed survival (Figure ??) did not differ between male and female patients (Hazard ratio for being male = 0.80, 95% confidence interval = [0.50, 1.29], p -value = 0.35) with there being no violation of the proportional hazards assumption ($\rho = 0.060$, $\chi^2 = 0.32$, $p = 0.58$).

3.3 Medication Differences – NEEDS REVISING

An omnibus MANOVA with mean daily doses for all 8 medications as dependent variables (Figure ??) and group (treatment resistant and responsive patient groups; TRS, NTR respectively) showed no significant differences ($F(1,143)=0.83$, Pillai's trace = 0.046, $p = 0.58$).

3.4 Symptom Domains – NEEDS REVISING

According to the TRRIP consensus, patients should further be categorised into treatment-resistance status according to whether positive, negative or cognitive symptoms dominate.

Positive and Negative Cases – NEEDS REVISING

Of the 86 TRS patients, none were exclusively treatment resistant in positive, negative or general domains alone, rather they were generally resistant on two or more domains. The percentage of treatment resistance patients were 63.95% (55) in all (positive, negative and general) domains; 2.32% (2) in positive and negative domains; 16.3% (14) in positive and general domains and 17.44% (15) in negative and general domains.

Cognitive Profiles – NEEDS REVISING

Of the 86 TRS patients, 16 were missing cognitive testing data in one of the five domains reported for the CATIE trial, rendering the overall composite cognitive test score missing. For the NTR patients, 5 of the 59 patients were similarly missing. The distribution of change in composite cognitive score was skewed, and violated assumptions of normality. A two-sample Kolmogorov-Smirnov test failed to reject the null hypothesis ($D = 0.19, p = 0.22$) that change in composite cognitive performance in the TRS and NTR groups are from the same distribution, suggesting there were no meaningful differences.

Burden of Individual Criteria – NEEDS REVISING

Figure ?? and Table 1 describe the proportions of patients who meet criteria on symptoms and SOF at the end of each phase in the CATIE trial design. Of note, there is a consistently high proportion of patients meeting the SOF criteria (moderate impairment) at each phase (which does not change with treatment).

| Phase | SOF | Symptoms | Adequate Rx |
|------------|-------------|------------|-------------|
| Baseline | 1320 (1445) | 757 (1206) | |
| Phase 1/1A | 1034 (1171) | 478 (1204) | 550 (1352) |
| Phase 1B | 95 (106) | 36 (107) | 51 (111) |
| Phase 2 | 413 (458) | 209 (484) | 200 (504) |
| Phase 3 | 207 (234) | 79 (249) | 111 (257) |

Table 1: Number of patients meeting threshold for each TRRIP criteria (total number with data available in parentheses). The ‘Adequate Rx’ column indicates the number of participants who had a single adequate trial in each phase of the trial

A total of 589 patients had exactly one adequate treatment episode across any phase of the trial. Of these, 26% (156) and 87% (503; excluding 5 patients missing sufficient data) were above threshold on the symptom or SOF criteria for TRS respectively. This sub-population represents 10% and 35% of the CATIE population for difficult-to-treat symptoms and moderate impairment on SOF respectively. Of the 156 patients with above threshold symptoms, 146 were simultaneously above threshold on SOF. The crude incidence for those meeting symptom criteria (despite one adequate trial) was 22.6 per 100 person-years and for those meeting SOF threshold, 74.5 per 100 person-years. For those meeting both symptoms and SOF criteria, the crude incidence was 21.21 per 100 person years.

The probability of leaving the trial meeting criteria on SOF, symptoms or both was estimated using logistic regression, with the duration of the adequate trial (in months), total number of other trials (i.e. those treatments trialled that did *not* meet criteria for dosing, duration etc.) and sex as predictors. Longer duration (in months) of the single adequate trial was associated with a modest reduction in the probability of meeting TRS criteria for symptoms

alone with odds ratios (OR) = 0.93 (p -value = 0.0003, 95% confidence interval = [0.90, 0.97]) and for both symptoms and SOF; OR = 0.94 (p -value = 0.0007, [0.90, 0.97]). However, no predictor was significant for SOF alone (full models reported in Supplementary Information, Table S3). In contrast, the same logistic regression model applied to the TRS group demonstrated *no* effect of adequate treatment duration, number of other treatment trials or sex (see Supplementary Information, Table S2).

Predicting TRS from Baseline Clinical Characteristics – NEEDS REVISING

Applying logistic regression to the 39 predictors for 1036 participants (that included 69 TRS patients) for which there was no missing data at baseline revealed modest associations for: verbal memory (OR = 0.60, 95% CI = [0.43, 0.84], p = 0.0031), positive symptom burden (OR = 1.07, 95% CI = [1.00, 1.14], p = 0.0407), age (OR = 1.05, 95% CI = [1.02, 1.09], p = 0.0048), being on ziprasidone at the start of the trial (OR = 2.63, 95% CI = [0.96, 6.51], p = 0.0454) or another antipsychotic *not* used in the CATIE trial (OR = 2.41, 95% CI = [1.11, 4.93], p = 0.0202). There were no other associations at statistical significance (Supplementary Information, Table S4).

Applying the ML procedure to the same data demonstrated poor predictive performance on out-of-sample tests, irrespective of whether SMOTE was used to balance classes during training, with the mean AUC = 0.56 (95% CI = [0.54, 0.59]), suggesting the signal in the inferential associations above are too weak to have utility for prediction.

4 Discussion – NEEDS REVISING

We demonstrate a first application of the TRRIP consensus by re-purposing the CATIE trial data and showed that when retrospective assessment was used to “screen-out” treatment resistance (i.e. according to the CATIE trial protocol),

cases were still present in the sample and emerged when prospective evaluation was applied as treatment progressed. We demonstrated that the incidence in at-risk patients (i.e. those having two or more adequate treatment trials) concords with the higher end of estimates in the literature. In addition, we demonstrated a significant burden of symptoms and/or impairment in social and occupational functioning in patients who had one adequate trial, suggesting a high level of unmet need in this “difficult to treat” subgroup and where the duration of that treatment was modestly associated with improvement in symptom (but not social and occupational functioning) criteria. The same association was not found in the TRS group, which provides support for their being a qualitative difference between those that will eventually respond to treatment, compared with those that are resistant. We showed that there was no differences in dosing between the TRS and NTR groups – suggesting that treatment resistance is not explicable by class or dose of medication alone – and there was no effect of sex on the hazard ratio for developing TRS.

In terms of predicting TRS from baseline clinical assessment data, we showed modest associations in a conventional inferential analyses which suggest that better verbal memory is associated with a slight reduction in odds ratio for developing TRS, and similarly, for age (with older patients being marginally more likely to meet TRS criteria) and PANSS positive symptoms (again, a small effect of higher symptom load with greater odds ratio for meeting TRS criteria). The effects of baseline drug treatment suggested the use of ziprasidone or another antipsychotic - not used in the CATIE study (risperidone olanzapine, quetiapine or perphenazine) had an increased risk of developing TRS. This needs to be prefaced by the caveat that these associations have wide confidence intervals, suggestive of estimation error secondary to small samples but may also reflect exposure to many antipsychotic medications prior to enrollment in the trial.

Despite these effects from inferential analyses, ML-based prediction of TRS status from baseline clinical data failed to show robust predictions in repeated out-of-sample cross validation, suggesting these have no utility for predicting

which patients will meet TRS using, at least given the assessment data in CATIE which includes the necessary variables for symptoms and SOF in the TRRIP consensus.

The CATIE trial excludes patients with first-episode psychosis so the population represents more chronic patients with median years since first prescribed an antipsychotic of 13 years (IQR = 18) and median age 42 (IQR = 16.5), so we would expect a higher level of treatment resistance. Of significance, using TRIPP, patient’s social and occupational impairment appears remarkably difficult to treat in both NTR, TRS and the subgroup we designated “difficult to treat” suggesting that – as a core feature of treatment resistance – social and occupational function is less discriminating in chronic patients.

Supplementary Information

S1.1 Current Symptoms and Response

The PANSS item scores were adjusted so that 0 represents an absence of symptoms, and consequently, 6 represents the most severe rating on any of the 30 items. This concurs with the methods described in (Leucht et al., 2009) and advocated in the TRRIP consensus for measuring symptom changes / response.

S1.2 Social and Occupational Function

The TRRIP consensus recommends using a validated scale such as the SOFAS, and sets the threshold for “moderate impairment” as a score of ≤ 60 on the 0-100 scale. The original SOFAS provides an unstructured and linguistically vague set of descriptors Morosini et al. (2000) whereas their Personal and Social Performance (PSP) scale provides an operationalised version with four domains:

- (A) socially useful activities, including work and study
- (B) personal and social relationships.
- (C) self-care
- (D) disturbing and aggressive behaviours

The equivalent of moderate impairment (on the SOFAS or Global Assessment of Function scale) using the PSP (Morosini et al., 2000) is defined as “Marked [difficulties] in 1 of general areas A-C, or manifest difficulties in D” which equates to a global assessment of functioning score of 51-60.

In the CATIE trial, there are three instruments which cover social and occupational functioning:

1. clinician-rated social functioning and interpersonal relationships are measured by the Heinrichs-Carpenter Quality of Life Scale (QLS) (Heinrichs et al., 1984)

2. the Lehman Quality of Life Interview (QOLI) measures physical, economic, social, and psychological functioning, employment, leisure, and residence Lehman (1988)
3. the MacArthur Community Violence Instrument (Steadman et al., 1998) is used to estimate risk of violence and aggression.

From these three instruments, we produced a proxy PSP score as follows:

- (A) socially useful activities, including work and study was measured by taking the “instrumental role” score on the Heinrichs-Carpenter QLS. A score of between 5-6 indicates “excellent / little impairment”, with a threshold of less than or equal to 3 indicating moderate impairment and 0-1 being severe impairment (Heinrichs et al., 1984).
- (B) personal and social relationships is equated with “interpersonal relations” score in the Heinrichs-Carpenter QLS, with the same scoring as “instrumental role” above
- (C) self-care is sub-divided into three domains, extracted from the QOLI and QLS:
 - (1) **accommodation status** where a maximum score of 5 indicates fully independent living, with the moderate impairment threshold at 3 “moderately supported accommodation”, 2 being “extremely restricted accommodation” and 1 equating to homelessness)
 - (2) **activities of daily living related to self-care** defined with 4 being the highest level (wash, do laundry, clean room, do chores) and a moderate threshold of 3 meaning professional help is required with at least one of these activities, and 0 being unable to do any of these tasks without help
 - (3) **independence** scored as 1 point for each of: reminding for a) appointments or b) to take medication and requiring supervision c) to take the patient to appointments or d) take medication – a score of

4 indicates complete independence, and a threshold of 3 was used for moderate impairment (i.e. at least one activity requires reminders with another requiring direct supervision).

The overall self-care (C) domain score is calculated to be at or below the “moderate impairment” threshold with a score of less than or equal to 3 in either (2) activities of daily living / (3) independence *or* less than or equal to 3 on (1) accommodation status, because living in moderately supported accommodation implies moderate impairment on the PSP Morosini et al. (2000) where they define “marked impairment” as “the person is still able to do something without professional or social help, although inadequately and/or occasionally; if helped by someone, he/she may be able to reach the previous level of functioning”.

(D) disturbing and aggressive behaviours are defined in the PSP (Morosini et al., 2000) as *severe* by “frequent verbal threats or frequent physical assaults, without intention or possibility to severe injuries” and suggests downgrading to “marked” impairment if occasional (rather than frequent) which is operationalised as less than 3 occurrences in a time period. We produced a proxy measure using the MacArthur scale by considering only items measuring aggression or violence directed at others by summing scores of 0 (did not occur) or 1 (did occur) over activities 1) throwing an object at someone, 2) pushed/grabbed/shoved someone, 3) slapped someone, 4) kicked/bitten/choked someone, 5) hit someone with fist/object or beaten up someone, 6) attempt to force someone to have sex, 7) threatened someone with gun/knife, 8) used knife/fired a gun on someone, 9) anything else considered violent, and 10) physically hurt someone – bruise, cut, broken bone, knocked unconscious – irrespective of mechanism). This yields a total score between 0-10. Given that *any* of these individual descriptors on the MacArthur scale would certainly meet the criteria for “severe” or “marked” impairment on the PSP, we set a threshold of greater than or

equal to 3 (by Morosini et al. (2000)’s definition of occasional) as being “marked” for the disturbing and aggressive behaviour domain in our proxy PSP measure.

After calculating the individual proxy domains (A) through (D) using the above rubric, we then set the overall SOF score to meet threshold for moderate impairment (i.e. meeting the TRRIP consensus for treatment resistance threshold on SOF) using the above rule: if the patient demonstrates “marked” impairment in one of A through C, or manifest difficulty in D. This equates to an overall 0-100 scaled score (e.g. for comparison with the SOFAS scale) of 51-60 and similarly, equates to a Global Assessment of Function (GAF) score indicating “moderate impairment” (Morosini et al., 2000). Of note, our rule for (D) cannot distinguish between “marked” and “manifest” (the lower rating) so it is likely our proxy SOFAS/PSP measure underestimates impairment on this domain, because only the more aggressive/violent (rather than disturbed) behaviours are captured. For missing data items, we assumed *no* impairment so again, our proxy for SOFAS is likely underestimating, rather than exaggerating, impairment in social and occupational functioning.

S1.3 Adequate Treatment

We sourced summary-of-product characteristics (SPCs) for each antipsychotic:

- Ziprasidone <http://labeling.pfizer.com/ShowLabeling.aspx?id=584>
- Olanzapine <http://pi.lilly.com/us/zyprexa-pi.pdf>
- Risperidone http://www.janssen.com/us/sites/www_janssen_com_usa/files/products-documents/risperdal-prescribing-information.pdf
- Perphenazine <https://www.medicines.org.uk/emc/medicine/22596>
- Quetiapine <https://www.medicines.org.uk/emc/medicine/2295>
- Aripiprazole https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/021436s038,021713s030,021729s022,021866s0231b1.pdf

| Medication | SPC Recommended Dose Range | | |
|--------------|----------------------------|--------------|---------------|
| | Minimum (mg) | Maximum (mg) | Midpoint (mg) |
| Ziprasidone | 40 | 200 | 120 |
| Olanzapine | 10 | 15 | 12.5 |
| Risperidone | 4 | 16 | 10 |
| Perphenazine | 12 | 24 | 18 |
| Quetiapine | 150 | 750 | 450 |
| Aripiprazole | 10 | 30 | 20 |
| Clozapine | 200 | 450 | 325 |
| Fluphenazine | 12.5 | 100 | 56.25 |

Table S1: Dose Thresholds for Adequate Treatment Trials

- Clozapine <https://www.medicines.org.uk/emc/medicine/32564>
- Fluphenazine long acting injectible/depot <http://www.medicines.org.uk/emc/medicine/6956/SPC/Modecate+Injection+25mg+m1>

The minimum adequate dose thresholds used are given in Table S1.3 and were determined as the mid-point between the minimum and maximum doses given in each medication’s SPC as: $\text{minimum dose} + (\text{maximum dose} - \text{minimum dose})/2$.

Only treatment episodes where the dose was greater than or equal to the midpoint were included as adequate treatment episodes.

S2 Secondary Analyses

S2.1 Burden of Individual TRS Criteria

For both the TRS group (having at least 2 adequate trials and meeting all criteria) and the difficult-to-treat group (meeting symptoms, SOF criteria (or both) but having had only 1 adequate trial) a logistic regression model was applied to estimate the effect of meeting TRS criteria with predictors: i) the number of other trials (beyond the adequate trials), ii) the duration (in months) of the adequate trial(s) and iii) sex.

For the TRS group (145 patients, 86 of which converted to TRS), there was no effect of these three predictors on symptoms, SOF or both (i.e. full criteria

for TRS) - Table S2.

| | Estimate | Std. Error | z-value | P-value | OR | Lower CI | Upper CI |
|------------------|----------|------------|---------|---------|------|----------|----------|
| Symptoms Only | | | | | | | |
| Rx Duration | -0.0520 | 0.0421 | -1.24 | 0.2166 | 0.95 | 0.87 | 1.03 |
| No. Other Rx | 0.0513 | 0.2737 | 0.19 | 0.8512 | 1.05 | 0.62 | 1.81 |
| Sex (male) | -0.5369 | 0.4256 | -1.26 | 0.2071 | 0.58 | 0.24 | 1.32 |
| SOF Only | | | | | | | |
| Rx Duration | 0.0251 | 0.1069 | 0.23 | 0.8144 | 1.03 | 0.83 | 1.28 |
| No. Other Rx | -1.1532 | 0.6830 | -1.69 | 0.0913 | 0.32 | 0.08 | 1.25 |
| Sex (male) | 0.9545 | 0.9622 | 0.99 | 0.3212 | 2.60 | 0.32 | 17.65 |
| SOF and Symptoms | | | | | | | |
| Rx Duration | -0.0568 | 0.0417 | -1.36 | 0.1736 | 0.94 | 0.87 | 1.02 |
| No. Other Rx | -0.0864 | 0.2707 | -0.32 | 0.7495 | 0.92 | 0.54 | 1.56 |
| Sex (male) | -0.4690 | 0.4166 | -1.13 | 0.2603 | 0.63 | 0.27 | 1.39 |

Table S2: Treatment Resistant (TRS) Group – SOF, Symptoms (PANSS) and both SOF and Symptoms in patients with two adequate treatment trial

For the difficult-to-treat group (589 patients, with 156 meeting symptom criteria, 513 meeting SOF criteria and 146 meeting both SOF and symptom criteria), there was a significant effect of duration of adequate trial for symptoms (but not SOF) and a model with the outcome being *both* symptoms and SOF (Table S3) with Bonferroni correction applied at the $0.05/3 = 0.016$

| | Estimate | Std. Error | z-value | P-value | OR | Lower CI | Upper CI |
|--------------------|----------|------------|---------|---------------|-------------|-------------|-------------|
| Symptoms Only | | | | | | | |
| Rx Duration | -0.0681 | 0.0190 | -3.59 | 0.0003 | 0.93 | 0.90 | 0.97 |
| No. Other Rx | -0.0715 | 0.1161 | -0.62 | 0.5378 | 0.93 | 0.74 | 1.17 |
| Sex (male) | 0.2557 | 0.2183 | 1.17 | 0.2416 | 1.29 | 0.85 | 2.00 |
| SOF Only | | | | | | | |
| Rx Duration | -0.0192 | 0.0250 | -0.77 | 0.4424 | 0.98 | 0.93 | 1.03 |
| No. Other Rx | 0.0868 | 0.1643 | 0.53 | 0.5973 | 1.09 | 0.80 | 1.52 |
| Sex (male) | 0.5204 | 0.2680 | 1.94 | 0.0522 | 1.68 | 0.98 | 2.83 |
| SOF and Symptoms | | | | | | | |
| Rx Duration | -0.0657 | 0.0194 | -3.39 | 0.0007 | 0.94 | 0.90 | 0.97 |
| No. Other Rx | -0.0410 | 0.1181 | -0.35 | 0.7285 | 0.96 | 0.76 | 1.21 |
| Sex (male) | 0.3799 | 0.2273 | 1.67 | 0.0947 | 1.46 | 0.94 | 2.31 |

Table S3: Difficult-to-treat Group : SOF, Symptoms (PANSS) and both SOF and Symptoms in patients with one adequate treatment trial with Bonferroni correction for significant results at $0.05 / 3$.

Predicting TRS Status from Baseline Characteristics

| | Estimate | Std. Error | z value | P-value | OR | Lower CI | Upper CI |
|---------------------|----------|------------|---------|---------------|---------------|---------------|---------------|
| SOF | | | | | | | |
| (A) Social | -0.0567 | 0.0949 | -0.5979 | 0.5499 | 0.9448 | 0.7806 | 1.1336 |
| (B) Personal | 0.1622 | 0.1180 | 1.3737 | 0.1695 | 1.1760 | 0.9314 | 1.4814 |
| (C) Accom | 0.0542 | 0.1279 | 0.4243 | 0.6713 | 1.0557 | 0.8278 | 1.3696 |
| (C) Self-care | -0.0470 | 0.1359 | -0.3456 | 0.7296 | 0.9541 | 0.7349 | 1.2540 |
| (C) Indepen | -0.0023 | 0.0941 | -0.0244 | 0.9805 | 0.9977 | 0.8309 | 1.2032 |
| (D) Aggression | -0.2169 | 0.2389 | -0.9079 | 0.3640 | 0.8050 | 0.4599 | 1.2028 |
| Cognitive Battery | | | | | | | |
| Verbal | -0.5027 | 0.1701 | -2.9562 | 0.0031 | 0.6049 | 0.4319 | 0.8420 |
| Vigilance | 0.1643 | 0.1685 | 0.9753 | 0.3294 | 1.1786 | 0.8484 | 1.6441 |
| Speed | 0.3928 | 0.1976 | 1.9878 | 0.0468 | 1.4811 | 1.0047 | 2.1831 |
| Reason | -0.1228 | 0.1683 | -0.7297 | 0.4655 | 0.8844 | 0.6364 | 1.2327 |
| Memory | -0.1202 | 0.1812 | -0.6635 | 0.5070 | 0.8867 | 0.6236 | 1.2717 |
| Symptoms (PANSS) | | | | | | | |
| Positive | 0.0690 | 0.0337 | 2.0466 | 0.0407 | 1.0714 | 1.0032 | 1.1452 |
| Negative | 0.0494 | 0.0283 | 1.7456 | 0.0809 | 1.0506 | 0.9938 | 1.1105 |
| General | -0.0014 | 0.0203 | -0.0674 | 0.9462 | 0.9986 | 0.9595 | 1.0392 |
| Demographics | | | | | | | |
| Age | 0.0492 | 0.0174 | 2.8199 | 0.0048 | 1.0504 | 1.0151 | 1.0871 |
| Sex | 0.1172 | 0.3264 | 0.3590 | 0.7196 | 1.1243 | 0.6030 | 2.1809 |
| White | 0.3783 | 0.9219 | 0.4104 | 0.6815 | 1.4598 | 0.2304 | 8.4625 |
| Black | 0.2272 | 0.9373 | 0.2424 | 0.8085 | 1.2550 | 0.1893 | 7.3538 |
| Native | 0.3311 | 1.1277 | 0.2936 | 0.7690 | 1.3926 | 0.0694 | 8.8562 |
| Asian | 0.8038 | 1.0488 | 0.7664 | 0.4435 | 2.2339 | 0.2680 | 15.5052 |
| Pacific | 0.2042 | 1.3906 | 0.1469 | 0.8832 | 1.2266 | 0.0415 | 13.2021 |
| Hispanic | -0.7623 | 0.5160 | -1.4773 | 0.1396 | 0.4666 | 0.1505 | 1.1800 |
| Illness | | | | | | | |
| Exacerbation | -0.0558 | 0.4046 | -0.1380 | 0.8903 | 0.9457 | 0.4200 | 2.0623 |
| CGI | -0.0995 | 0.1943 | -0.5121 | 0.6086 | 0.9053 | 0.6168 | 1.3222 |
| 1st Rx E+B | -0.0143 | 0.0217 | -0.6569 | 0.5112 | 0.9858 | 0.9423 | 1.0266 |
| 1st AP Rx | -0.0048 | 0.0236 | -0.2020 | 0.8399 | 0.9952 | 0.9520 | 1.0451 |
| Lifetime Admissions | -0.1216 | 0.1079 | -1.1261 | 0.2601 | 0.8855 | 0.7166 | 1.0954 |
| Admission past Yr | 0.2143 | 0.1794 | 1.1942 | 0.2324 | 1.2390 | 0.8564 | 1.7383 |
| Hx ASB | 0.0552 | 0.0893 | 0.6176 | 0.5368 | 1.0567 | 0.8819 | 1.2534 |
| TD | -0.7229 | 0.4185 | -1.7271 | 0.0841 | 0.4854 | 0.1998 | 1.0499 |
| Baseline Rx | | | | | | | |
| Olz | -0.5901 | 0.3649 | -1.6172 | 0.1058 | 0.5543 | 0.2615 | 1.1051 |
| Que | 0.0577 | 0.4241 | 0.1360 | 0.8918 | 1.0594 | 0.4321 | 2.3221 |
| Ris | -0.2759 | 0.3601 | -0.7662 | 0.4436 | 0.7589 | 0.3629 | 1.5031 |
| Zip | 0.9665 | 0.4830 | 2.0010 | 0.0454 | 2.6287 | 0.9574 | 6.5106 |
| Hal | 0.0524 | 0.5732 | 0.0914 | 0.9272 | 1.0538 | 0.2941 | 2.9352 |
| Depot | -12.9699 | 703.5753 | -0.0184 | 0.9853 | NE | NE | NE |
| Per | -0.3738 | 1.0885 | -0.3434 | 0.7313 | 0.6881 | 0.0359 | 3.9319 |
| Other | 0.8778 | 0.3779 | 2.3227 | 0.0202 | 2.4056 | 1.1093 | 4.9327 |

Table S4: Logistic Regression (1036 patients, of which 69 were TRS patients) for TRS status given baseline data. Bold variables are statistically significant at $p < 0.05$. Lower/Upper CI for 95% confidence. NE indicates CI inestimable (due to low numbers). SOF - social and occupational functioning scores computed as described in section S1.2 above. Demographics - sex, ethnicity and racial heritage characteristics are 0/1 coded. Illness: Exacerbation - crisis stabilisation / hospital treatment in past 3 months (0/1); CGI - clinical global impression (severity); 1st Rx E+B - years since first treatment for emotional/behavioural difficulties; 1st AP Rx - years since first treatment with antipsychotic medication; Lifetime Admissions - lifetime number of hospital admissions; Admissions past Yr - number of hospital admissions in past year; Hx ASB - history of childhood antisocial behaviour; TD - meets criteria for tardive dyskinesia. Baseline Rx : 0/1 indicating baseline use of Olz (Olanzapine) Que (Quetiapine).

References

- Blagus, R. and Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14(1):106.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.
- Dammak, M. (2013). Treatment-resistant schizophrenia: Prevalence and risk factors in Woolfolk, R. (Ed.), *Mental Disorders - Theoretical and Empirical Perspectives*, InTech, DOI: 10.5772/52430. Available from: <https://www.intechopen.com/books/mental-disorders-theoretical-and-empirical-perspectives/treatment-resistant-schizophrenia-prevalence-and-risk-factors>.
- Elkis, H. (2007). Treatment-resistant schizophrenia. *Psychiatric Clinics of North America*, 30(3):511–533.
- Essock, S. M., Hargreaves, W. A., Dohm, F.-A., Goethe, J., Carver, L., and Hipshman, L. (1996). Clozapine eligibility among state hospital patients. *Schizophrenia bulletin*, 22(1):15–25.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1).
- He, H. and Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.
- Heinrichs, D. W., Hanlon, T. E., and Carpenter Jr, W. T. (1984). The quality of life scale: an instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bulletin*, 10(3):388–398.
- Howes, O. D., McCutcheon, R., Agid, O., de Bartolomeis, A., van Beveren, N. J., Birnbaum, M. L., Bloomfield, M. A., Bressan, R. A., Buchanan, R. W.,

- Carpenter, W. T., Castle, D. J., Citrome, L., Daskalakis, Z. J., Davidson, M., Drake, R. J., Dursun, S., Ebdrup, B. H., Elkis, H., Falkai, P., Fleischacker, W. W., Gadelha, A., Gaughran, F., Glenthøj, B. Y., Graff-Guerrero, A., Hallak, J. E., Honer, W. G., Kennedy, J., Kinon, B. J., Lawrie, S. M., Lee, J., Leweke, F. M., MacCabe, J. H., McNabb, C. B., Meltzer, H., Möller, H.-J., Nakajima, S., Pantelis, C., Reis Marques, T., Remington, G., Rossell, S. L., Russell, B. R., Siu, C. O., Suzuki, T., Sommer, I. E., Taylor, D., Thomas, N., Üçok, A., Umbricht, D., Walters, J. T., Kane, J., and Correll, C. U. (2017). Treatment-Resistant Schizophrenia: Treatment Response and Resistance in Psychosis (TRRIP) Working Group Consensus Guidelines on Diagnosis and Terminology. *American Journal of Psychiatry*, 174(3):216–229.
- Juarez-Reyes, M., Shumway, M., Battle, C., Bacchetti, P., Hansen, M., and Hargreaves, W. (1995). Restricting clozapine use: the impact of stringent eligibility criteria. *Psychiatric Serv*, 46:801–6.
- Keefe, R. S., Bilder, R. M., Harvey, P. D., Davis, S. M., Palmer, B. W., Gold, J. M., Meltzer, H. Y., Green, M. F., Miller, D. D., Canive, J. M., et al. (2006). Baseline neurocognitive deficits in the catie schizophrenia trial. *Neuropsychopharmacology*, 31(9):2033.
- Lehman, A. F. (1988). A quality of life interview for the chronically mentally ill. *Evaluation and program planning*, 11(1):51–62.
- Lehman, A. F., Lieberman, J. A., Dixon, L. B., McGlashan, T. H., Miller, A. L., Perkins, D. O., Kreyenbuhl, J., American Psychiatric Association, J. S., Steering Committee on Practice Guidelines, S. C., Altshuler, K., Cook, I., Cross, C. D., Mellman, L., Moench, L. A., Norquist, G., Twemlow, S. W., Woods, S., Yager, J., Gray, S. H., Askland, K., Pandya, R., Prasad, K., Johnston, R., Nininger, J., Peele, R., Anzia, D. J., Benson, R. S., Lurie, L., Walker, R. D., Kunkle, R., Simpson, A., Fochtmann, L. J., Hart, C., and Regier, D. (2004). Practice guideline for the treatment of patients with

- schizophrenia, second edition. *The American journal of psychiatry*, 161(2 Suppl):1–56.
- Leucht, S., Davis, J. M., Engel, R. R., Kissling, W., and Kane, J. M. (2009). Definitions of response and remission in schizophrenia: recommendations for their use and their presentation. *Acta Psychiatrica Scandinavica*, 119(438):7–14.
- Leucht, S., Samara, M., Heres, S., and Davis, J. M. (2016). Dose Equivalents for Antipsychotic Drugs: The DDD Method. *Schizophrenia Bulletin*, 42(suppl 1):S90–S94.
- Lieberman, J. A., Stroup, T. S., McEvoy, J. P., Swartz, M. S., Rosenheck, R. A., Perkins, D. O., Keefe, R. S., Davis, S. M., Davis, C. E., Lebowitz, B. D., Severe, J., and Hsiao, J. K. (2005). Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *New England Journal of Medicine*, 353(12):1209–1223.
- Meltzer, H. and Kostakoglu, A. (2001). Treatment resistant schizophrenia in J. Lieberman and R. Murray (Eds), (2012) *Comprehensive Care of Schizophrenia: A Textbook of Clinical Management*, Oxford University Press.
- Morosini, P.-L., Magliano, L., Brambilla, L., Ugolini, S., and Pioli, R. (2000). Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. *Acta Psychiatrica Scandinavica*, 101:323–329.
- Steadman, H. J., Mulvey, E. P., Monahan, J., Robbins, P. C., Appelbaum, P. S., Grisso, T., Roth, L. H., and Silver, E. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of general psychiatry*, 55(5):393–401.
- Szymanski, S., Lieberman, J. A., Alvir, J. M., Mayerhoff, D., Loebel, A., Geisler, S., Chakos, M., Koreen, A., Jody, D., Kane, J., et al. (1995). Gender differences in onset of illness, treatment response, course, and biologic

indexes in first-episode schizophrenic patients. *American Journal of Psychiatry*, 152(5):698–703.