

Matheus Rodrigues Luiz

**ANÁLISE DO ESCORE POLIGÊNICO DE RISCO COMO PREDITOR
DE RESPOSTA À RISPERIDONA: aplicação em uma coorte de
primeiro episódio psicótico**

Trabalho de Conclusão de Curso
apresentado à Universidade Federal de
São Paulo – Escola Paulista de Medicina,
para obtenção do título de Tecnólogo em
Informática em Saúde

São Paulo

2022

Matheus Rodrigues Luiz

**ANÁLISE DO ESCORE POLIGÊNICO DE RISCO COMO PREDITOR
DE RESPOSTA À RISPERIDONA: Aplicação em uma coorte de
primeiro episódio psicótico**

Trabalho de Conclusão de Curso
apresentado à Universidade Federal de
São Paulo – Escola Paulista de Medicina,
para obtenção do título de Tecnólogo em
Informática em Saúde

Orientadora:

Profa. Dra. Sintia Iole Nogueira Belangero

São Paulo

2022

Ficha catalográfica

Luiz, Matheus Rodrigues

Análise do Escore Poligênico de Risco como preditor de resposta à risperidona: Aplicação em uma coorte de primeiro episódio psicótico /

Matheus Rodrigues Luiz. - São Paulo, 2022. XII, 21f.

Trabalho de Conclusão de Curso - Universidade Federal de São Paulo, Escola Paulista de Medicina. Curso Superior de Tecnologia em Informática em Saúde.

Título em inglês: Analysis of the Polygenic Risk Score as a predictor for risperidone treatment response: application in a first-episode of psychosis cohort.

1. Bioinformática. 2. Psiquiatria. 3. Genética Médica. 4. Farmacogenética. 5. Transtornos Psicóticos.

UNIVERSIDADE FEDERAL DE SÃO PAULO
ESCOLA PAULISTA DE MEDICINA
CURSO SUPERIOR DE TECNOLOGIA EM INFORMÁTICA EM SAÚDE

Coordenadora do curso:

Profa. Dra. Viviane Bernardo

**ATA DE DEFESA PÚBLICA DA UNIDADE CURRICULAR TRABALHO DE
CONCLUSÃO DE CURSO**

**Curso Superior de Tecnologia em Informática Em Saúde
Escola Paulista de Medicina - Universidade Federal de São Paulo**

Em 05 de agosto de 2022, em sessão de apresentação remota, constituiu-se a banca examinadora do trabalho intitulado "**Análise do Escore Poligênico de Risco como preditor de resposta à risperidona: aplicação em uma coorte de primeiro episódio psicótico**" da UC Trabalho de Conclusão de Curso do(a) discente **Matheus Rodrigues Luiz**, matriculado(a) no Curso Superior de Tecnologia em Informática em Saúde, composta pela professora orientadora Dra. Sintia Iole N. Belangero e pelos membros convidados Dra. Letícia Maria N. Spíndola e Prof. Dr. Paulo Bandiera Paiva, sendo presidida pela professora orientadora Dra. Sintia Iole N. Belangero.

Após reunião em sessão reservada, a banca examinadora deliberou e decidiu pela (aprovação/reprovação) do referido trabalho com a nota final

.....

Parecer do Orientador:

O aluno teve excelente desempenho durante toda a trajetória de execução do seu trabalho de conclusão de curso, atingindo todas as metas, além de mostrar atuação acima da esperada para um aluno de graduação.

Aluno aprovado com nota 10.

Comentários da Banca:

Letícia: O aluno apresentou a aula de forma clara e com o domínio do assunto. A monografia está bem escrita, demonstrando o domínio do projeto pelo aluno. Minha única sugestão em relação a monografia é a correção da interpretação dos Odds Ratio nos resultados. Considero o aluno aprovado.

Signature: Leticia Spindola
Leticia Spindola (Aug 5, 2022 16:22 GMT+2)

Email: leticia.n.spindola@gmail.com

Signature: PBandieraPaiva
PBandieraPaiva (Aug 5, 2022 11:55 ADT)

Email: paiva@unifesp.br

Nada mais tendo a constar, assinam esta ata, professora orientadora e componentes da banca examinadora.

Assinatura: 

Profa. Dra. Sintia Iole Nogueira Belangero, orientadora

Profa. Adjunta do Departamento de Morfologia e Genética da UNIFESP/EPM

Assinatura: Leticia Spindola

Dra. Letícia Maria Nery Spíndola

Postdoctoral fellow na Universidade de Bergen, Noruega

Assinatura: P.Bandiera-Paiva

Prof. Dr. Paulo Bandiera Paiva

Prof. Associado do Departamento de Informática em Saúde da UNIFESP/EPM

Data: 05 de agosto de 2022

Dedicatória

*À minha mãe, minha maior apoiadora em
todos os meus sonhos e projetos,
desde sempre e para sempre.*

Agradecimentos

Agradeço, primeiramente, **a Deus**, pela força e ajuda concedida para chegar até aqui.

Ao **meu pai, minha mãe e meu irmão**, por serem simplesmente incríveis e terem me dado todo apoio desde o início da graduação, e serem a minha retaguarda para todos os momentos.

À **minha família**, minha base, por todo amor, apoio e carinho em todos esses anos de universidade.

À **minha prima Débora**, minha dupla desde o início da vida, hoje minha conselheira com quem compartilho todos os altos e baixos da vida adulta

À **minha orientadora, Profa. Dra. Sintia Belangero**, que me fez brilhar os olhos para a área de genética e me ensinou não só o que diz respeito à linha de pesquisa, mas muito sobre a vida, e me acolheu nos meus piores momentos.

Ao **grupo de pesquisa Psychiatric Genetics Team**, desde seus ex-integrantes até os atuais, que influenciaram (e muito) tudo o que sei a respeito de genética, e me proporcionaram os melhores momentos que tive na graduação.

Aos **meus coorientadores da iniciação científica**, Dra. Vanessa Ota e Giovany Costa, que foram meu auxílio pra toda as horas em todas as minhas dúvidas e, com certeza, contribuíram muito para o meu aprendizado.

Ao **meu namorado, Matheus Citibaldi**, por me lembrar todos os dias como a vida é boa ao lado de quem te quer bem, e pelos incentivos diários (e recíprocos) na carreira científica. Obrigado por não desistir dos nossos sonhos.

À minha professora de biologia do ensino médio, **Aldenice**, que mesmo não gostando de genética, plantou uma semente que hoje gerou este trabalho.

Aos meus amigos que sempre estiveram comigo, **Iago, Muri, Isa e Giovanna**, a vida seria muito triste sem o apoio de vocês e sem vocês por perto.

À **turma de Tecnologias em Saúde TS11**, em especial à **Bia, Bruna, Gio, Graci, Gabi Minohara, Gabi Caldeira, Judô, Ju, Neto, Sté e Vini**, por simplesmente serem o melhor grupo que um graduando poderia pedir e pelo carinho e amizade que permanecem até hoje

Ao **Dan, a Jess, Adri, Amanda, Bia, Joice, Ju, Rafa, Liri e Gabe**, por compartilharem o dia a dia da genética e do LiNC comigo nos últimos anos e fazer do laboratório um ambiente muito agradável pra se desenvolver pesquisa.

À **Júlia Tezzei**, pelo braço forte quando eu mais precisei.

À **Cintia Scala**, por ser uma mentora tão atenciosa, tanto no âmbito profissional quanto pessoal, e compartilhar dos meus melhores momentos na Roche.

À **Ana Cristina**, minha antiga dupla de análise de dados na Roche, pelo crescimento exponencial técnico que você me possibilitou desenvolver

À **Paula Becker**, pelo apoio em todos os âmbitos da vida, desde o artístico até o pessoal.

A **todos os docentes e funcionários** da Escola Paulista de Medicina / Universidade Federal de São Paulo, que contribuíram para meu desenvolvimento pessoal e acadêmico, em especial os professores do Departamento de Informática em Saúde, da Disciplina de Genética e do Departamento de Psiquiatria.

Às **agências de fomento FAPESP, CAPES e CNPq** que possibilitaram o desenvolvimento deste projeto.

À **banca**, por ter aceitado avaliar este trabalho.

E a todos que contribuíram para minha formação pessoal e profissional, bem como para o desenvolvimento deste trabalho.

Epígrafe

*“Quando sentir que a esperança se foi,
Olhe dentro de si e seja forte (...)*

Existe um herói em você”

Mariah Carey

Resumo

Objetivo: analisar o escore poligênico de risco para esquizofrenia como possível variável preditora da resposta ao tratamento com risperidona, baseando-se em dois diferentes critérios de resposta.

Método: Foram selecionados 99 pacientes em primeiro episódio psicótico, previamente virgens de tratamento com antipsicótico que foram avaliados por meio da escala PANSS (Escala de Sintomas Positivos e Negativos em Esquizofrenia) em dois momentos, no início do tratamento com antipsicóticos e após dois meses de acompanhamento. A partir desse acompanhamento foi possível calcular a taxa de resposta ao tratamento e classificar os pacientes em respondedor e não respondedor baseado em dois critérios de resposta (redução de 30% e 50% da escala PANSS). Os pacientes tiveram seu sangue coletado para realizar a genotipagem, utilizada para calcular o escore poligênico de risco para esquizofrenia que foi analisado como variável preditora do desfecho resposta, junto a outras covariáveis, por meio de uma regressão logística.

Resultado: os modelos de predição mostraram que pacientes com maior escore poligênico de risco para esquizofrenia tem maior chance de não responder ao tratamento, para os dois critérios de resposta. Os melhores modelos de predição, para os dois critérios de resposta, envolveram a covariável tempo de psicose não tratada, mas em ambos os critérios, nenhuma variável foi suficiente para predizer a resposta ao tratamento sozinha e o melhor modelo em ambos os critérios envolvia todas as variáveis. Houve uma tendência à significância estatística em muitos achados, evidenciando a necessidade de se replicar este estudo em um número amostral maior.

Conclusão: pacientes com maior escore poligênico para esquizofrenia demonstraram maior risco para não responderem ao tratamento com antipsicóticos, e o escore poligênico de risco se mostrou um bom preditor de resposta quando atrelado a variáveis clínicas como tempo de psicose não tratada, histórico familiar, idade e sexo biológico.

Abstract

Objective: Analyze the polygenic risk score for schizophrenia as a possible predictor variable for treatment response with risperidone, based in two different response criteria

Methods: We recruited 99 patients in First-Episode of Psychosis, previously antipsychotic-naive, who were evaluated by the PANSS scale in two different moments: in the beginning of the treatment with antipsychotics (baseline) and a follow-up after two months in treatment. This follow-up made possible to calculate de response rate and to classify patients in treatment responders and non-responders (based in two different criteria – 30% and 50% reduction of PANSS score). Patients had their blood collected for genotyping and the genotype data was used to generate the polygenic risk score for schizophrenia, which was analyzed as a predictor for the response outcome, along with the covariables, in a logistic regression.

Results: The prediction models show that patients with higher PRS for schizophrenia have higher chance of not responding to treatment, in both response criteria. The best prediction models had the covariable Duration of Untreated Psychosis but, in both criteria, a single variable, alone, was not efficient in predicting the response outcome. Therefore, the best prediction model in both criteria was the model with all the independent variables. Results show a tendency to statistical significance, indicating the need to replicate this study in a bigger sample size.

Conclusion: Patients with higher PRS for schizophrenia showed higher risk for not responding to the treatment with antipsychotics. The PRS was a good predictor for response in a model with another clinical variables as predictors, such as duration of untreated psychosis, familiar history, age and biological sex.

Sumário

Dedicatória.....	vii
Agradecimentos.....	viii
Epígrafe.....	x
Resumo.....	xi
Abstract.....	xii
Lista de abreviaturas, siglas e símbolos.....	xiv
1 INTRODUÇÃO.....	1
2 Objetivos.....	4
3 métodos.....	5
3.1 Delineamento do Estudo.....	5
3.2 Coorte de Primeiro Episódio Psicótico.....	5
3.2.1 Critérios de inclusão e exclusão.....	6
3.2.2 Critérios de resposta e composição dos grupos.....	6
3.3 Material biológico.....	7
3.4 Genotipagem.....	7
3.5 Escore Poligênico de Risco.....	7
3.5.1 Controle de Qualidade.....	7
3.5.2 Cálculo de escore poligênico de risco.....	8
3.7 Análises Estatísticas.....	10
4 RESULTADOS.....	11
4.1 Análise Descritiva da Amostra.....	11
4.2 Escore Poligênico de Risco.....	14
4.3 Modelos de Regressão Logística.....	16
5 discussão.....	18
6 CONCLUSÃO.....	21
7 REFERÊNCIAS.....	22

Lista de abreviaturas, siglas e símbolos

AIC	<i>Akaike Information Criterion</i> (Critério de Informação de Akaike)
DUP	<i>Duration of Untreated Psychosis</i> (Tempo de Psicose Não-Tratada)
GWAS	<i>Genome-Wide Association Study</i> (Estudo de associação genética em larga-escala ou escala genômica)
IC	Intervalo de Confiança
PEP	Primeiro Episódio Psicótico
PRS	<i>Polygenic Risk Score</i> (Escore Poligênico de Risco)
SNP	<i>Single Nucleotide Polymorphism</i> (Polimorfismo de Nucleotídeo Único)
TCLE	Termo de Consentimento Livre e Esclarecido

1 INTRODUÇÃO

A psicose é uma condição grave que resulta na dificuldade de se determinar o que é real, como nas alucinações e delírios, podendo ser dividida em três grupos: psicose idiopática, psicose devido às condições de saúde (como transtornos neurodegenerativos) e psicose tóxica (devido ao abuso de substâncias)¹. Os transtornos que apresentam esse sintoma, denominados transtornos psicóticos, têm prevalência de 4.03 a cada 1000 indivíduos, em 12 meses². Dentre os diagnósticos, encontram-se a esquizofrenia³, o transtorno afetivo bipolar⁴ e o transtorno depressivo maior⁵.

A esquizofrenia é uma doença que afeta 1,5 em cada 10.000 habitantes no mundo por ano⁶ e é uma das principais causas de perda de anos de vida saudável⁷. Ela é caracterizada pela psicose, que compõe o quadro de sintomas positivos, e pelo embotamento afetivo, alogia e avolição, que compõem os sintomas negativos. A presença de sintomas cognitivos como déficit de atenção, de memória e de funções executivas também são características dessa doença.⁸

A esquizofrenia é caracterizada como multifatorial e poligênica, pois envolve variantes comuns em múltiplos genes e estes atuam em conjunto com fatores ambientais, aumentando o risco para desenvolvê-la⁹. Assim, diversos fatores — como os genéticos, o uso de drogas, as complicações obstétricas, entre outros¹⁰ — podem contribuir para o desenvolvimento de transtornos psicóticos. Isso os classifica como doenças multifatoriais, isto é, apresentam uma etiologia complexa com contribuição de fatores genéticos e ambientais. A combinação desses fatores é heterogênea, o que confere ao transtorno um fenótipo também heterogêneo, quando comparadas diferenças fenotípicas interindividuais¹¹.

O primeiro episódio psicótico (PEP) geralmente ocorre entre o final da adolescência e início da vida adulta, momento em que o indivíduo está no auge do seu potencial produtivo, levando a um grande prejuízo funcional e na qualidade de vida.¹²

Após o PEP, o paciente é submetido a um tratamento com antipsicóticos. As taxas de resposta a antipsicóticos podem variar de 17 a 81% em um período de 16

semanas, mostrando que ainda existe uma parcela substancial de pacientes que não respondem ao tratamento¹³. Dentre os antipsicóticos mais utilizados em PEP está a risperidona, um fármaco bloqueador de receptores de dopamina e de serotonina¹⁴, classificada como um antipsicótico de segunda geração (ou atípico).

As diferentes respostas individuais a um determinado tratamento, bem como a presença de efeitos adversos a ele é estudado por meio da Farmacogenética, uma área da Genética que estuda a influência das variantes genéticas sobre o metabolismo — geral e de fármacos específicos — de um indivíduo. Atualmente, tem sido uma área muito estudada com o crescimento da medicina de precisão¹⁵.

Como supramencionado, entre os fatores que podem contribuir para as diferenças interindividuais de resposta ao tratamento estão as variantes genômicas. Dependendo da combinação de variantes genéticas, o indivíduo pode metabolizar a risperidona de forma lenta ou rápida^{16,17}, levando a concentrações diferentes dos níveis da droga e/ou seus metabólitos no sangue e alterando a resposta clínica.

Polimorfismos de nucleotídeo único (*single nucleotide polymorphisms* - SNPs) são uma das principais variações genéticas estudadas pela farmacogenética. Esses polimorfismos são alterações em uma única base no genoma que ocorrem com frequência maior que 1% na população¹⁸.

Essas variantes podem ser analisadas por meio de estudos de associação, como os Estudos de Associação Genômica em Larga Escala (GWAS – *Genome-Wide Association Study*). O GWAS é um método capaz de analisar milhares de SNPs espalhados pelo genoma e comparar a frequência alélica dessas variantes genéticas entre dois grupos, geralmente caso e controle, e associá-las a um determinado fenótipo¹⁹. Esse tipo de estudo é realizado não somente na área de psiquiatria, mas também em outras áreas com diversos fenótipos, tais como, o câncer de mama, diabetes e doenças do sono²⁰.

GWAS presentes na literatura utilizaram amostras, predominantemente, de ancestralidade europeia. De acordo com dados do *GWAS Catalog*, aproximadamente 79% de todos os indivíduos participantes de GWAS têm ancestralidade europeia²¹. Isso pode conduzir a um viés em relação à população europeia, e dificulta o desenvolvimento de ferramentas de predição de risco que sejam replicáveis em diferentes populações devido à diferença nas frequências alélicas entre elas²². Esse

efeito confundidor é especialmente significativo em amostras muito miscigenadas, como a brasileira^{23,24}. Estudos prévios mostraram que a inclusão de amostras não europeias aumenta a acurácia de predição^{25,26}.

Além de identificar novos alvos e genes específicos para a doença, o GWAS demonstrou que nenhum gene é suficiente para explicar a doença e muitos genes atuam para o desenvolvimento do fenótipo. Neste contexto, Purcell et al., (2009) criaram um modelo poligênico levando em consideração milhares de variantes para pontuar riscos individuais²⁷, chamado de Escore Poligênico de Risco (PRS, do inglês *Polygenic Risk Score*). Em linhas gerais, utilizando-se das estatísticas resumidas de uma amostra base de GWAS para determinado fenótipo, é possível calcular, em uma outra amostra independente, um escore de risco individual para o mesmo fenótipo.

Assim, o PRS está baseado na soma dos alelos de risco dos SNPs analisados no GWAS, ponderados pelo seu tamanho de efeito, conferindo um determinado risco individual. Pode ser usado para inferir a arquitetura genética de um traço específico e, também, agir como biomarcador para um fenótipo.²⁸ O escore já foi utilizado para prever o risco de se desenvolver doenças como a esquizofrenia. Um estudo caso-controle concluiu que o efeito cumulativo de múltiplos alelos associados a esquizofrenia explicam uma variância de até 24%, apesar do baixo tamanho de efeito individual de um alelo²⁹.

Espera-se que o PRS, em uma amostra, adote uma distribuição conforme a figura abaixo (fig. 1), no qual os pacientes apresentam uma maior média do escore (representando um maior risco genético) em relação aos controles.

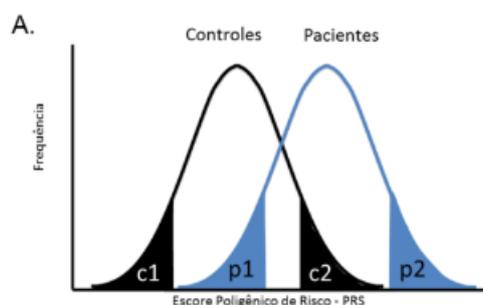


Figura 1.: os indivíduos localizados em c1 e p1 apresentam um menor PRS, comparado aos indivíduos localizados em c2 e p2. A média do escore dos pacientes é maior que a dos controles (adaptado de Talarico et. al, 2018)³⁰

2 OBJETIVOS

O presente estudo teve como objetivo geral analisar o escore poligênico de risco para esquizofrenia como possível variável preditora da resposta ao tratamento com risperidona, baseando-se em dois diferentes critérios de resposta.

Os objetivos específicos desse estudo foram:

1) analisar a variância explicada obtida a partir do escore poligênico de risco para esquizofrenia em cada um dos critérios de resposta.

2) comparar diferentes modelos de regressão logística utilizando o escore poligênico de risco como preditor para diferentes taxas de resposta ao tratamento.

3 MÉTODOS

3.1 Delineamento do Estudo

Este foi um estudo de coorte prospectivo longitudinal que avaliou 99 pacientes em primeiro episódio psicótico e virgens de antipsicótico em dois momentos: 1) antes do tratamento com risperidona (*baseline*) e 2) dois meses após o tratamento padronizado com risperidona (seguimento). Os pacientes foram avaliados (por meio de escalas psiquiátricas e sociodemográficas) nos dois momentos e submetidos à coleta de sangue (para posterior extração de DNA) no *baseline*.

3.2 Coorte de Primeiro Episódio Psicótico

Essa coorte é formada por sujeitos em Primeiro Episódio Psicótico com idade entre 15 e 42, sendo todos virgens de tratamento com antipsicótico, provenientes do Centro de Atenção Integrada à Saúde Mental (CAISM – Vila Mariana, São Paulo) da Escola Paulista de Medicina da Universidade Federal de São Paulo (EPM/UNIFESP).

Todos os indivíduos ou seus responsáveis legais assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), aprovado pelo Comitê de Ética em Pesquisa da UNIFESP (processo 0603/10).

Na linha de base (*baseline*), os pacientes foram avaliados por meio da Entrevista Clínica Estruturada para DSM-IV (SCID-I). Outros instrumentos foram aplicados para avaliar os sintomas e resposta farmacológica (como descrito a seguir), mas para o presente projeto, apenas a PANSS foi utilizada.

1. PANSS – Escala de Sintomas Positivos e Negativos³¹
2. CGI - Impressão de Gravidade Clínica³²
3. GAF –Avaliação Global do Funcionamento³³
4. CDSS – Escala Calgary de Depressão para Esquizofrenia

Um acompanhamento do paciente, após dois meses da primeira avaliação, foi realizado com os mesmos instrumentos aplicados na etapa de inclusão, para verificar a taxa de resposta ao tratamento e confirmar o diagnóstico.

3.2.1 Critérios de inclusão e exclusão

Para serem incluídos nesse estudo, os indivíduos deveriam apresentar diagnóstico de primeiro episódio psicótico (DSM-IV), e não ter recebido tratamento com antipsicótico anteriormente

Também só foram incluídos indivíduos que tivessem concluído a avaliação com a escala PANSS nos dois momentos, a fim de avaliar a resposta ao tratamento.

Foram excluídos os indivíduos com histórico de convulsões, tumor encefálico, e deficiência intelectual.

3.2.2 Critérios de resposta e composição dos grupos

A resposta ao tratamento foi definida de acordo com base na pontuação da escala PANSS. A PANSS é dividida em 3 domínios: PANSS positiva (7 itens), PANSS negativa (7 itens) e de psicopatologia geral (16 itens). Para cada item, o paciente pode pontuar de 1 (leve) a 7 (grave). Para calcular a taxa de resposta, primeiramente, foram subtraídos 30 pontos da PANSS total (pontuação mínima), obtendo-se o valor bruto da PANSS total para ambos os momentos. Em seguida, a $PANSS_1$ (no primeiro momento) é subtraída da $PANSS_2$ (no segundo momento) e a resposta ao tratamento será calculada como uma porcentagem da $PANSS_1$ pela seguinte fórmula:

$$\%Resposta = 100 \cdot \frac{PANSS_1 - PANSS_2}{PANSS_1}$$

Assim, se um paciente atingiu a pontuação de 150 no primeiro momento e reduziu para 90, ele respondeu 40%.

Os pacientes foram classificados em dois grupos de acordo com a resposta ao tratamento com a risperidona: respondedor ou não-respondedor, para cada uma das taxas de corte (*cut-offs*) a seguir: redução de 30% e 50%³¹ do escore total da PANSS. Assim, utilizamos dois grupos para cada desfecho como a seguir.

- Resposta de 30%: respondedor (resposta \geq 30%), com 62 pacientes x não respondedor (<30%), com 32 pacientes
- Resposta de 50%: respondedor (resposta \geq 50%) x não respondedor (<50%), com 47 pacientes em ambos os grupos

3.3 Material biológico

Foram coletados 10 mL de sangue periférico, com material estéril e descartável, para extrações de DNA e RNA, além de outros estudos envolvendo a coorte. Neste projeto, foram utilizados apenas os dados provenientes do DNA extraído a partir de tubos EDTA pelo kit *Gentra Puregene* (Qiagen-Sciences, USA). As amostras de DNA extraídas foram quantificadas pelo aparelho NanoDrop® e analisadas pelas razões entre os comprimentos de onda 260/230 e 260/280³⁴. Posteriormente, uma alíquota do DNA extraído foi utilizada para a genotipagem (detalhada a seguir).

3.4 Genotipagem

A genotipagem das amostras de DNA da coorte de pacientes em PEP foi realizada por meio de uma parceria da Disciplina de Genética do Departamento de Morfologia e Genética da UNIFESP com o *Institute of Psychiatry do King's College London* (London, UK) utilizando o *Illumina PsychArray*, que é capaz de avaliar aproximadamente 593 mil marcadores genéticos.

3.5 Escore Poligênico de Risco

O escore poligênico de risco (PRS) é um valor estimado da susceptibilidade genética de um indivíduo a um fenótipo. Seu cálculo se dá a partir da soma dos alelos de risco daquele indivíduo para o fenótipo escolhido, ponderados pelo tamanho de efeito da variável, estimado a partir de um GWAS de referência, ou seja, estudo previamente realizado, para o fenótipo escolhido³⁵.

As etapas para cálculo do escore poligênico de risco estão descritas a seguir:

3.5.1 Controle de Qualidade

Inicialmente, foi realizada a etapa de controle de qualidade por meio do software Plink nos dados de genotipagem, a fim de excluir SNPs e indivíduos que apresentam valores abaixo dos parâmetros estabelecidos (tabela 1) e que foram definidos com base no Consórcio de Genômica Psiquiátrica (*Psychiatric Genomics Consortium* – PGC). Os dados de genotipagem passaram pela seguinte filtragem:

- a. Remoção de SNPs com falha na genotipagem em mais de 10% da amostra
- b. Remoção de indivíduos com falha na genotipagem em mais de 10% dos SNPs
- c. Remoção de SNPs fora do Equilíbrio de Hardy-Weinberg
- d. Remoção de indels e SNPs cuja frequência do alelo raro é menor que 1%
- e. Indivíduos com alto grau de parentesco (medido pelo parâmetro PIHAT).

3.5.2 Cálculo de escore poligênico de risco

Para o cálculo do Escore Poligênico de Risco (PRS), foi utilizada a ferramenta PRSice-2³⁶, criada para automatizar as análises de PRS. O PRSice é um programa de linha de comando que estende os comandos em R usados na ferramenta Plink.

Para calcular o PRS, primeiramente foi utilizado a estatística resumida (que contém a identificação do SNP, o valor de p e tamanho de efeito para cada SNP do GWAS) de um GWAS (com grande tamanho amostral) com fenótipo de interesse, o qual é nomeada amostra de base, e que nosso caso, escolhemos o estudo do consórcio de genômica psiquiátrica (PGC)³⁷, focando nos estudos de esquizofrenia fase 3 (PGC-SCZ3), que contém o maior N amostral do mundo até o momento (~76 mil indivíduos com diagnóstico de esquizofrenia e ~243 mil indivíduos controle)²⁹.

Dessa forma, os SNPs foram ordenados em uma lista de acordo com seu respectivo valor de p , e um escore para cada SNP foi calculado, com base na quantidade de alelos de risco presente para cada SNP, ponderadas pelo seus respectivos tamanho de efeito. Os escores de cada SNP são somados gerando um PRS único para aquele indivíduo e para aquele fenótipo escolhido.

Em uma etapa anterior ao cálculo do PRS propriamente dito, realizamos uma etapa para verificar qual o melhor modelo para o seu cálculo. Essa escolha é realizada com base na variância explicada para cada modelo (aquele que apresenta a maior variância explicada, ou seja, que explica a maior parte da variância do fenótipo de estudo, será utilizado). Entre os modelos de PRS testados, testamos diferentes p -thresholds (P_t) para inclusão ou não de SNPs com base em seus valores de p provenientes das estatísticas resumidas da amostra base (GWAS anterior): esse P_t pode variar de $P_t < 5 \times 10^{-8}$ (ou seja, só as variantes que são

estatisticamente significantes no GWAS após a correção para múltiplas comparações) até $P_t=1.0$ (ou seja, todas as variantes testadas).

Em seguida, a partir da soma de número de alelos de risco que o indivíduo possui: um (heterozigoto para o alelo de risco), dois (homozigoto para o alelo de risco) ou nenhum alelo de risco (homozigoto para o alelo selvagem), ponderados pelos seus respectivos tamanhos de efeito (ou *odds ratio*) obtidos no resultado do GWAS base, um escore individual é calculado (Purcell et al. 2009; Dudbridge 2013). A Figura 2a resume o cálculo do PRS.

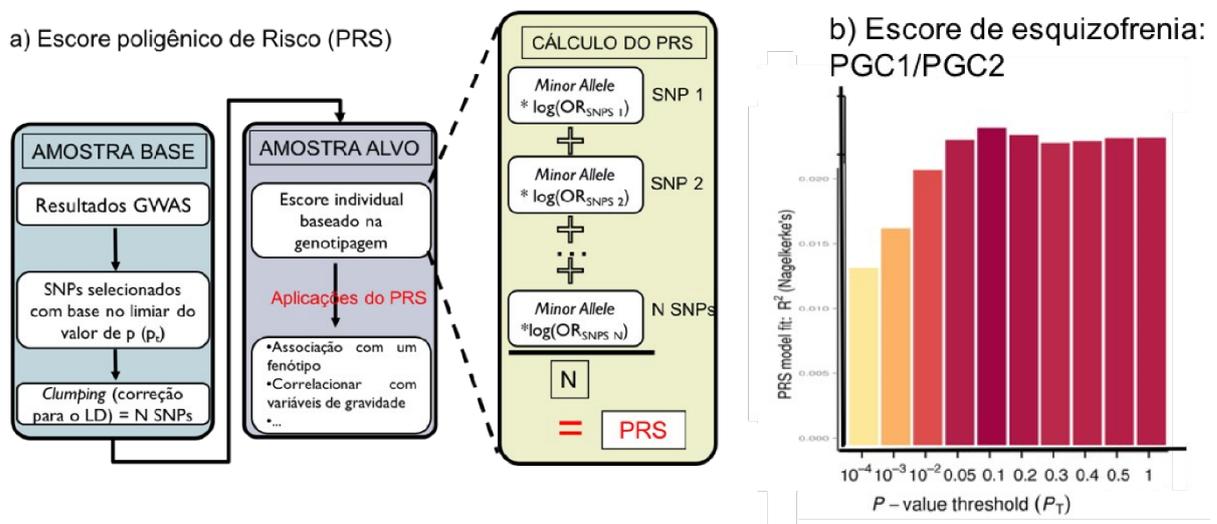


Figura 2. Escore poligênico de risco (PRS). A) Esquema simplificado do cálculo do PRS. B) Resultados do PRS de esquizofrenia de uma amostra base (PGC2) e amostra alvo (PGC1) do Psychiatric Genomics Consortium (PGC). A variância explicada pela correlação de Nagelkerke entre casos e controles foi de 30% para P_t maiores que 0.01, além disso, a variância explicada aumentou com o aumento do P_t e se manteve constante após o P_t de 0,1.

3.7 Análises Estatísticas

Modelos de regressão logística foram construídos para cada uma das duas variáveis dependentes (variáveis preditas): resposta ao tratamento de 30% e resposta ao tratamento de 50%.

Em todos os modelos, o PRS foi utilizado como variável independente (preditora) e foram acrescentados ao modelo diferentes covariáveis tais como: tempo de psicose não tratada (DUP), histórico de psicose na família, idade e sexo biológico.

No banco de dados da coorte, as variáveis estão divididas da seguinte forma:

Tabela x: disposição das variáveis clínicas da coorte

Variável:	Formato
Idade	Número inteiro
Sexo	Categórico: "Masculino" "Feminino"
Histórico Familiar	Categórico: "Sim" "Não"
Tempo de Psicose Não Tratada	Categórico: "Menos de 1 mês" "1-6 meses" "6-12 meses" "mais de um ano"

Também foram calculados 10 componentes principais da genotipagem, também por meio do software Plink, a fim de utilizá-los como covariáveis dos modelos de regressão e evitar vieses de estratificação populacional.

Para todas as análises, o índice de significância adotado foi de 0,05.

4 RESULTADOS

4.1 Análise Descritiva da Amostra

Foram incluídos no estudo 99 pacientes de uma coorte de Primeiro Episódio Psicótico que possuíam tanto dados da escala PANSS quanto dados de genotipagem. A partir dos filtros de controle de qualidade, foram removidos pacientes e variantes genéticas antes do cálculo do escore poligênico, conforme tabela 1:

Tabela 1: Número de pacientes e variantes excluídas conforme critérios de controle de qualidade

	Parâmetros analisados	Valores dos parâmetros	Nº de SNPs excluídos	Nº de indivíduos excluídos
SNPs	Frequência do alelo raro	<1% (0,01)	50993	-
	Baixa taxa de genotipagem	<1% (0,01)	806	-
	Equilíbrio de Hardy-Weinberg	$P < 5 \times 10^{-6}$	895	-
Indivíduos	Baixa taxa de genotipagem	<1% (0,01)	-	0
	PIHAT	> 0,12	-	0
	Indivíduos sem informação de PANSS em 2º momento			5
	Total		335.134	94

Os grupos respondedores X não-respondedores foram formados a partir desses 94 pacientes, conforme quadro abaixo:

Quadro 1: Classificação de pacientes conforme critério de resposta

Grupo	30%	50%
Respondedores	62	47
Não Respondedores	32	47

Para a criação dos modelos com variáveis clínicas, foram selecionados apenas os pacientes que continham as informações das covariáveis Idade, sexo, tempo de

psicose não tratada e histórico familiar, formando uma subamostra de 77 pacientes, utilizada em todos os modelos de regressão logística. As variáveis idade e sexo biológico estavam presentes em todos os pacientes da subamostra.

Tabela 2: Análise descritiva da variável DUP. Número de pacientes em cada categoria.

	Tempo de Psicose Não Tratada			
	Critério de 30%		Critério de 50%	
	Respondedores	Não Respondedores	Respondedores	Não Respondedores
Menos de 1 mês	25	6	22	9
1-6 meses	16	8	13	11
6-12 meses	6	6	2	10
mais de um ano	4	6	3	7

Tabela 3: Análise descritiva da variável HF. Número de pacientes em cada categoria.

	Histórico de Psicose na Família			
	Critério de 30%		Critério de 50%	
	Respondedores	Não Respondedores	Respondedores	Não Respondedores
Sim	29	11	26	23
Não	12	15	14	14

Tabela 4: Análise descritiva da variável Sexo Biológico. Número de pacientes em cada categoria.

	Sexo Biológico			
	Critério de 30%		Critério de 50%	
	Respondedores	Não Respondedores	Respondedores	Não Respondedores
Feminino	24	8	19	13
Masculino	27	18	21	24

Tabela 5: Análise descritiva da variável Idade.

	Idade			
	Critério de 30%		Critério de 50%	
	Respondedores	Não Respondedores	Respondedores	Não Respondedores
Média	24,3	27,4	24	26,7
Mediana	21	26	21,5	26
Interquartile Range	11	10	9.5	13

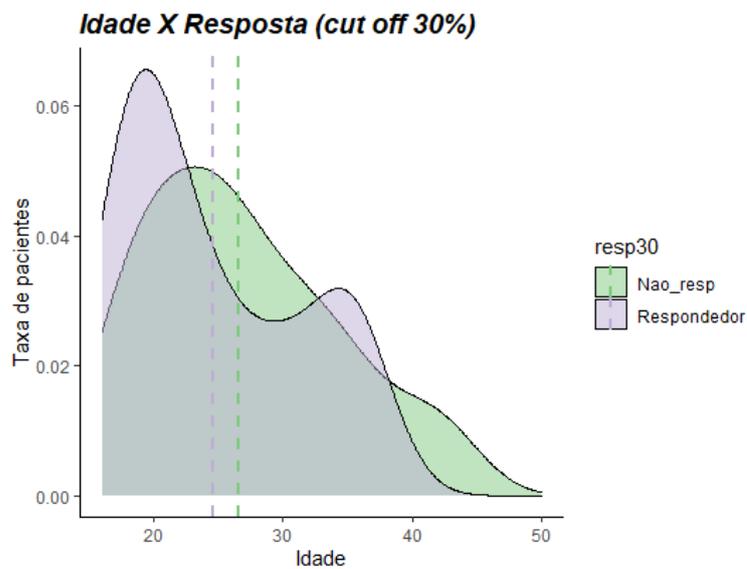


Figura 3: curva de distribuição da variável idade para os grupos respondedor e não respondedor, utilizando o critério de 30%

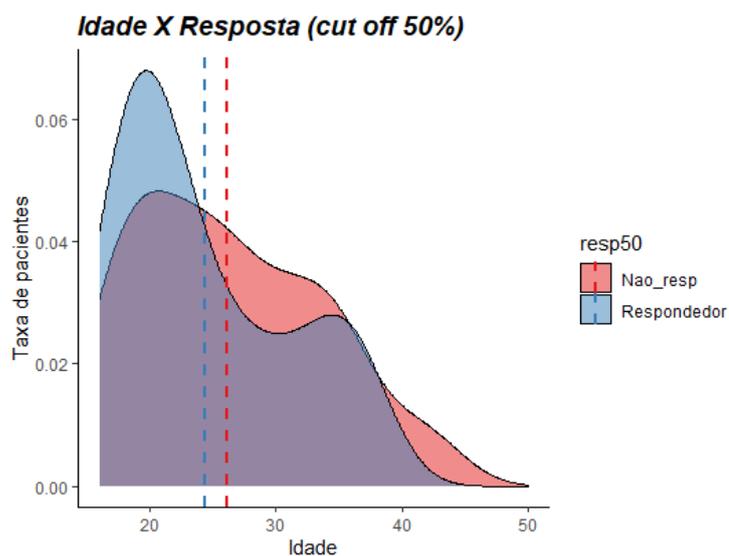
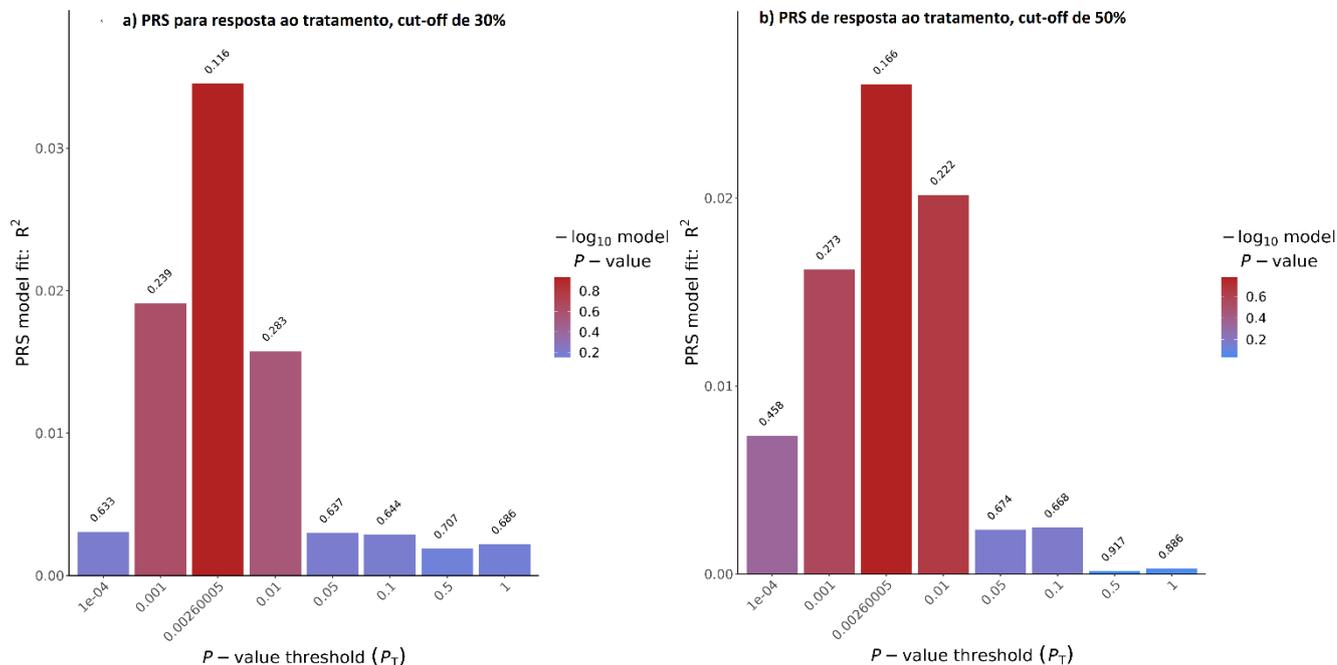


Figura 4: curva de distribuição da variável idade para os grupos respondedor e não respondedor, utilizando o critério de 50%

4.2 Escore Poligênico de Risco

Calculamos o escore para dois grupos de resposta, 30% e 50%, conforme figura 5 abaixo



Fig(5). A) e B) gráficos em barra evidenciando o melhor *threshold* de variantes que explicam o fenótipo de resposta ao tratamento. No eixo X encontram-se os *thresholds* (de 0.0001 a 1) e no eixo Y, a variância explicada. Os valores sobre as barras indicam o valor de p de uma regressão logística, gerada pela ferramenta PRSice, relacionando cada PRS para dado *threshold* ao fenótipo.

O melhor *p-threshold*, tanto para 30%, quanto para 50% de resposta, foi o de $p < 2,6 \cdot 10^{-3}$, com 4208 SNPs incluídos nos modelos. Entretanto, a variância explicada diferiu entre esses fenótipos, mostrando variância explicada de 3,2% para a resposta de 30% e 2,5% para o fenótipo de resposta de 50%.

Partindo do cálculo do escore poligênico, foi analisada a distribuição deste entre respondedor X não-respondedor, para ambos os grupos

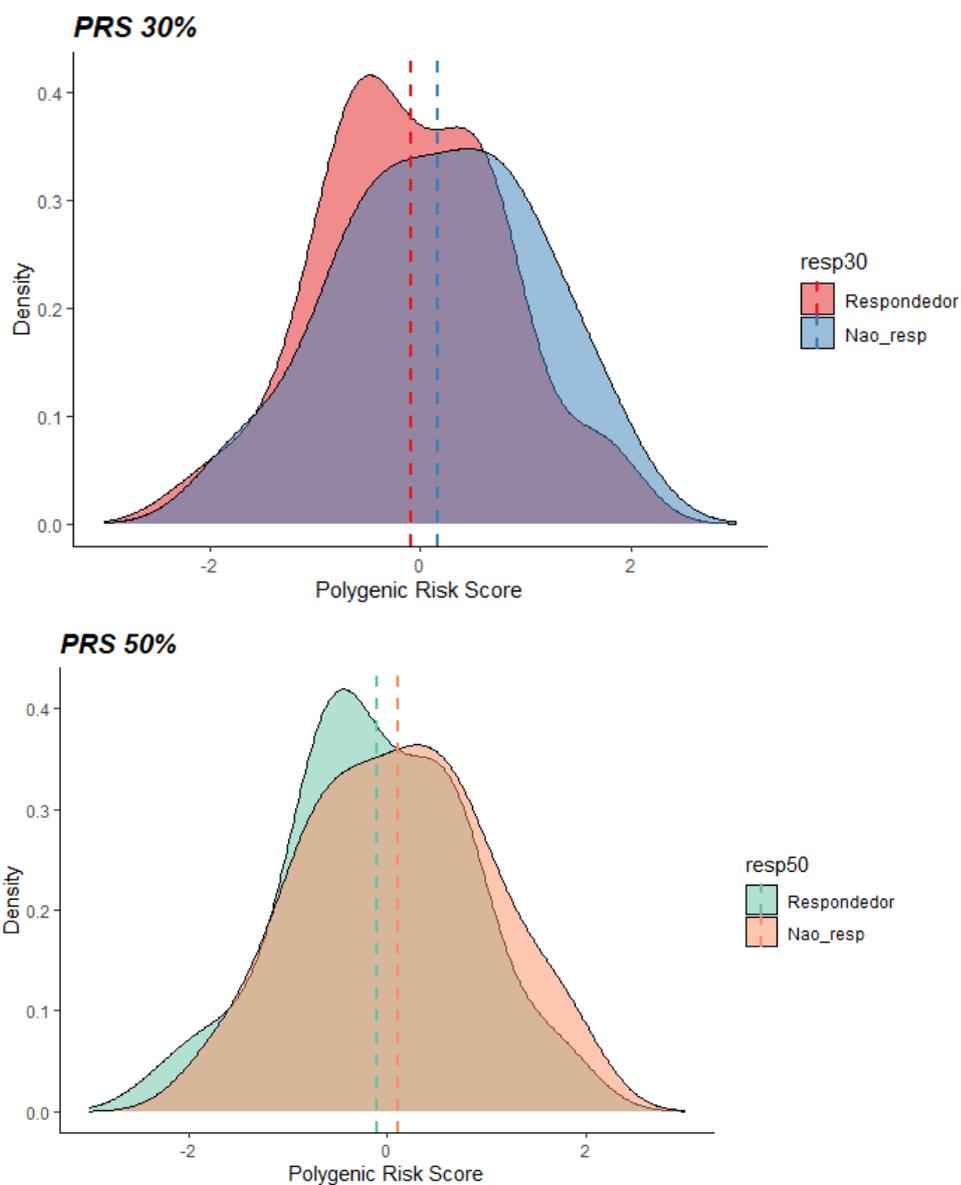


Fig 6 A) e B) evidenciam a distribuição dos escores poligênicos entre os fenótipos Respondedor ao Tratamento e Não respondedor, para os critérios 30% de resposta e 50%. A média dos sujeitos não respondedores, em ambos os grupos, mostrou-se maior do que a média dos sujeitos respondedores.

4.3 Modelos de Regressão Logística

Para os dois critérios de resposta, foram criados os mesmos modelos de regressão logística, com as covariáveis: tempo de psicose não tratada (DUP), histórico familiar (HF), sexo e idade. Para todos os modelos, a variável dependente (predita) será o desfecho de resposta ao tratamento, enquanto o escore poligênico e demais covariáveis serão variáveis independentes e preditoras. Abaixo, resumimos os modelos e suas variáveis.

- 1) PRS, covariando para os 10 componentes principais (modelo base)
- 2) Modelo Base + DUP
- 3) Modelo base + HF
- 4) Modelo base + DUP + HF
- 5) Modelo base + DUP + Sexo + Idade
- 6) Modelo Base + Sexo + Idade
- 7) Todas as variáveis

O escore poligênico, então, foi usado como principal variável preditora do desfecho resposta ao tratamento (variável predita). O desempenho desses modelos e suas variáveis estão descritos conforme tabelas abaixo.

Fenótipo de 30% de resposta ao tratamento

Modelo	R ² Nagelkerke	AIC	P-Value PRS	Odds Ratio PRS	IC PRS 2.5%	IC PRS 97.5%
PRS	0.169	112.41	0.067	1.46	0.27	0.82
PRS + DUP	0.321	108.12	0.128	1.57	0.87	2.84
PRS + HF	0.188	113.24	0.176	1.48	0.83	2.63
PRS + DUP + HF	0.329	109.27	0.131	1.57	0.87	2.86
PRS + DUP + Sexo + idade	0.431	103.27	0.067	1.83	0.95	3.49
PRS + Sexo + Idade	0.31	106.76	0.121	1.60	0.0007	0.18
Todas as Variáveis	0.456	103.72	0.062	1.87	0.96	3.64

Modelo	p-valor das covariáveis			
	DUP	HF	Sexo	Idade
PRS + DUP	0.029*	-	-	-
PRS + HF	-	0.282	-	-
PRS + DUP + HF	0.036*	0.457	-	-
PRS + DUP + Sexo + idade	0.055	-	0.047*	0.041*
Todas as Variáveis	0.071	0.170	0.030*	0.028*

Tabela 6: Desempenho das variáveis e modelos de regressão para o critério de 30% de resposta ao tratamento. Variáveis: DUP (Tempo de Psicose Não Tratada), HF (Histórico Familiar), Sexo Biológico e Idade.

Com base no R^2 Nagelkerke e AIC, os dois modelos com melhor desempenho foram modelos 5 e 7, que incluíam as variáveis: 5) DUP, sexo biológico e idade e 7) com todas as variáveis testadas. Esses também foram os modelos onde o PRS apresentou o melhor desempenho como preditor (menor valor de p e maior *odds ratio*).

Fenótipo de 50% de resposta ao tratamento

Modelo	R^2 Nagelkerke	AIC	P-Value PRS	Odds Ratio PRS	IC PRS 2.5%	IC PRS 97.5%
PRS	0.11	122.35	0.86	1.30	0.78	2.19
PRS + DUP	0.33	113.54	0.20	1.43	0.82	2.50
PRS + HF	0.23	116.31	0.31	1.32	0.76	2.27
PRS + DUP + HF	0.42	108.39	0.22	1.43	0.80	2.56
PRS + DUP + Sexo + idade	0.41	111.46	0.09	1.71	0.92	3.19
PRS + Sexo + Idade	0.23	118.88	0.15	1.49	0.85	2.63
Todas as Variáveis	0.53	102.49	0.077	1.85	0.93	3.68

Modelo	p-valor das covariáveis			
	DUP	HF	Sexo	Idade
PRS + DUP	0.012*	-	-	-
PRS + HF	-	0.006*	-	-
PRS + DUP + HF	0.017*	0.012*	-	-
PRS + DUP + Sexo + idade	0.015*	-	0.022*	0.096
Todas as Variáveis	0.03*	0.0004*	0.02*	0.03*

Tabela 7: Desempenho das variáveis e modelos de regressão para o critério de 50% de resposta ao tratamento. Variáveis: DUP (Tempo de Psicose Não Tratada), HF (Histórico Familiar), Sexo Biológico e Idade

Para o critério de 50% de resposta ao tratamento, os dois melhores modelos (maior número de Nagelkerke e menor número de AIC) foram os 4 e 7, sendo: 4) com as variáveis tempo de psicose não tratada (DUP) e histórico familiar (HF) e 7) com todas as variáveis, respectivamente.

Entretanto, para este fenótipo, o PRS não mostrou seu melhor desempenho nos modelos de regressão supramencionados, sendo sua melhor performance encontrada em modelo similar ao usado para resposta de 30%: modelo com todas as variáveis e modelo com as covariáveis DUP, sexo biológico e idade.

5 DISCUSSÃO

Para esse estudo, utilizamos uma coorte de pacientes em primeiro episódio psicótico e virgens de tratamento com antipsicóticos, o que torna nosso estudo único e com grandes vantagens sobre outros da literatura. Com esse estudo temos a possibilidade de evitar fatores confundidores, como alguns presentes em outros estudos, tais como tempo/progressão da doença e efeito de tratamento sobre o fenótipo. Outra vantagem é estarmos estudando uma coorte com tratamento padronizado com um único antipsicótico (risperidona), evitando assim, diferentes interferências e nos levando a uma maior confiabilidade sobre o fenótipo estudado.

Não há na literatura um consenso concreto do que seria resposta ao tratamento baseando-se na escala PANSS. Entretanto, encontram-se artigos que consideram vários cut-offs. Os critérios de 30% e 50%³¹ foram escolhidos para este estudo de forma a avaliar a predição do PRS levando em consideração um critério mais brando de resposta e um mais rígido.

Em todos os modelos de regressão logística realizados, os valores de p para o Escore Poligênico de Risco não foram estatisticamente significativos. O que pode ser explicado de duas formas: a) há uma necessidade de se aumentar o tamanho amostral e b) reduzir o viés de ancestralidade.

É importante notar que, mesmo com esses vieses, a distribuição dos PRSs nos dois grupos de resposta, 30% e 50%, adotaram a forma esperada, onde a média dos não-respondedores é maior que a média dos respondedores³⁰ Como o GWAS base não foi um GWAS de resposta ao tratamento, e sim de esquizofrenia, era esperado que a média fosse maior nos não-respondedores. Entretanto, é visível no gráfico alguns picos que deformam a curva do que seria considerada uma distribuição normal. Esses picos são causados pelos vieses de ancestralidade.

Os resultados mostram uma tendência à significância estatística, mesmo com número amostral pequeno. Essa tendência mostra o potencial do estudo nessa coorte, pelo fenótipo limpo, de pacientes sem uso prévio de antipsicóticos, que permite analisar os desfechos sem interferência de fatores confundidores.

O escore poligênico de risco tem sido uma ferramenta muito estudada na área de genômica. Entretanto, nas amostras brasileiras, ainda há uma dificuldade de utilizá-lo, tendo em vista que os GWAS de referência têm, em sua maioria, amostras europeias e, ao analisarmos uma amostra miscigenada baseada em uma amostra maior de ancestralidade definida, pode ocorrer o viés de ancestralidade, o que não permite ver o real poder de predição e variância explicada do escore poligênico.

Com o aumento exponencial do tamanho da amostra dos consórcios internacionais, esta ferramenta torna-se ainda mais robusta. Uma vantagem do PRS é que não se faz necessária a utilização de grandes N amostrais desde que seu cálculo seja baseado em um grande GWAS³². O potencial do PRS é muito amplo, e nos próximos anos pode tornar-se um passo fundamental para qualquer estudo que envolva o cálculo de risco genético para alguma doença poligênica.

Na área de psiquiatria, o PRS tem apresentado alguns resultados interessantes, como a associação entre os oito principais transtornos psiquiátricos³⁸. Para a esquizofrenia, o PRS tem sido correlacionado com as variáveis quantitativas, como a gravidade dos sintomas (Derks et al. 2012) e da atividade pré-frontal³⁹. Para outros transtornos psiquiátricos, o PRS foi correlacionado com dados de neuroimagem funcional e estrutural (transtorno bipolar)^{40,41} e com a redução do volume cortical em regiões específicas (transtorno depressivo)⁴².

Embora um dos primeiros estudos a utilizar PRS tenha sido com esquizofrenia, atualmente, o PRS vêm sendo aplicado com sucesso para outros fenótipos não-psiquiátricos, como diabetes, câncer, estatura entre outros. Entretanto, há poucos estudos na literatura^{43,44} que o relacione à resposta ao tratamento com antipsicóticos, evidenciando a necessidade de se estudar e replicar estes achados, o que também pode contribuir, futuramente, para a viabilização de um tratamento precoce para a psicose, visto que há na literatura evidência que piores prognósticos estão relacionados com maior tempo entre o primeiro episódio de psicose e o início do tratamento^{45,46}. Dessa forma, a descoberta de um escore genético que possa auxiliar a prever resposta ao tratamento poderá reduzir o tempo de tratamento inadequado, melhorando o prognóstico do paciente.

Corroborando com estes achados, a variável que mais influenciou nos modelos de regressão logística foi o Tempo de Psicose Não Tratada (DUP), isto é, o tempo entre o início dos sintomas e o início do tratamento com antipsicóticos. Isso corrobora com o achado da literatura que discute a DUP não como fator da gravidade da doença no momento baseline e sim nas diferenças de resposta, sendo um potencial fator ambiental a ser estudado para medir seus reais efeitos e correlacionar com fatores genéticos. A inclusão dessa variável nos modelos tornou a predição destes melhorada em ambos os critérios de resposta.

Estes critérios foram escolhidos pois não há um consenso do que é resposta ao tratamento baseado na escala PANSS. Mas, sabe-se que uma porcentagem específica pode ser considerada um critério mais brando enquanto outra um critério mais rígido. Nesse estudo testamos os dois cut-offs a fim de verificar não só as diferenças quanto à taxa de remissão dos sintomas e as variáveis clínicas e genéticas envolvidas, quanto ao tamanho amostral que, por ser pequeno, variou significativamente de um critério de resposta para outro, o que influenciou no resultado dos modelos.

A variável Histórico familiar, apesar de significativa no modelo com todas as variáveis para o critério de 50%, não aumentou muito o poder de predição (medido pelo R^2) dos modelos em que ela foi inserida. Essa variável foi analisada partindo da alta herdabilidade de transtornos mentais⁴⁷, em especial a esquizofrenia⁴⁸. Porém, a sua baixa performance pode se dar pela característica multifatorial da psicose, evidenciando a presença de fatores ambientais na etiologia do episódio psicótico.

Ainda que a inserção das variáveis supramencionadas tenha causado uma melhora significativa nos modelos, nenhuma delas sozinha se mostrou capaz de uma predição tão significativa sozinha, ou seja, os melhores modelos foram aqueles com todas as variáveis. Entretanto, a variável tempo de psicose não tratada se mostrou um potencial a ser estudada e replicada num maior tamanho amostral.

É importante considerar o tamanho amostral deste projeto e que, com uma replicação com um tamanho amostral, as variáveis podem ter um nível maior de significância estatística, bem como melhor performance do escore poligênico nos modelos.

6 CONCLUSÃO

A razão de chances do escore poligênico em todos os modelos deste trabalho se mostrou coerente à hipótese inicial de que indivíduos com maior escore poligênico para esquizofrenia tem maior risco de não responderem ao tratamento com risperidona. Dentre os dois critérios de resposta, o PRS teve melhor predição nos modelos que incluíam a variável Tempo de Psicose Não Tratada, evidenciando um bom modelo de predição para resposta ao tratamento que envolva variável genética e ambiental.

Contudo, vale ressaltar que mesmo o tempo de psicose não tratada sozinho não foi capaz de criar um modelo de predição tão bom junto ao escore poligênico quanto a inserção de todas as variáveis, para ambos os critérios de resposta.

Apesar do escore ter demonstrado boa performance nessa amostra, ainda está longe de poder ser aplicado na prática clínica, de forma a prever algum desfecho e influenciar na tomada de decisão ao tratamento. Mais estudos são necessários de forma a concretizar seu uso.

7 REFERÊNCIAS

1. Lieberman, J. A. & First, M. B. Psychotic Disorders. *New England Journal of Medicine* **379**, 270–280 (2018).
2. Moreno-Küstner, B., Martín, C. & Pastor, L. Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses. *PLoS One* **13**, e0195687 (2018).
3. Tandon, R. *et al.* Definition and description of schizophrenia in the DSM-5. *Schizophr. Res.* **150**, 3–10 (2013).
4. Anderson, I. M., Haddad, P. M. & Scott, J. Bipolar disorder. *BMJ* **345**, (2012).
5. Epidemiology of depression with psychotic experiences and its association with chronic physical conditions in 47 low- and middle-income countries | Psychological Medicine | Cambridge | Core.
<https://www.cambridge.org/core/journals/psychological-medicine/article/epidemiology-of-depression-with-psychotic-experiences-and-its-association-with-chronic-physical-conditions-in-47-low-and-middle-income-countries/447B12B6BD246AA3B57C97687F77A527> (2019).
6. McGrath, J. *et al.* A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med* **2**, 13 (2004).
7. Mathers, C. D., Lopez, A. D. & Murray, C. J. L. The Burden of Disease and Mortality by Condition: Data, Methods, and Results for 2001. in *Global Burden of Disease and Risk Factors* (eds. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J.) (World Bank, 2006).

8. Tamminga, H. M. I. and C. A. Treating Impaired Cognition in Schizophrenia. *Current Pharmaceutical Biotechnology* <http://www.eurekaselect.com/98885/article> (2012).
9. Smeland, O. B., Frei, O., Dale, A. M. & Andreassen, O. A. The polygenic architecture of schizophrenia - rethinking pathogenesis and nosology. *Nat Rev Neurol* **16**, 366–379 (2020).
10. Vilain, J. *et al.* [Environmental risk factors for schizophrenia: a review]. *Encephale* **39**, 19–28 (2013).
11. Kraguljac, N. V. *et al.* Neuroimaging Biomarkers in Schizophrenia. *Am J Psychiatry* **178**, 509–521 (2021).
12. Malla, A. & Payne, J. First-episode psychosis: psychopathology, quality of life, and functional outcome. *Schizophr Bull* **31**, 650–671 (2005).
13. Carbon, M. & Correll, C. U. Clinical predictors of therapeutic response to antipsychotics in schizophrenia. *Dialogues Clin Neurosci* **16**, 505–524 (2014).
14. Germann, D., Kurylo, N. & Han, F. Risperidone. *Profiles Drug Subst Excip Relat Methodol* **37**, 313–361 (2012).
15. Daly, A. K. Pharmacogenetics: a general review on progress to date. *Br. Med. Bull.* **124**, 65–79 (2017).
16. Llerena, A., Berez, R., Peñas-Lledó, E., Süveges, A. & Fariñas, H. Pharmacogenetics of clinical response to risperidone. *Pharmacogenomics* **14**, 177–194 (2013).
17. Mas, S. *et al.* Intuitive pharmacogenetic dosing of risperidone according to CYP2D6 phenotype extrapolated from genotype in a cohort of first episode psychosis patients. *Eur Neuropsychopharmacol* **27**, 647–656 (2017).

18. Shastry, B. S. SNPs: impact on gene function and phenotype. *Methods Mol. Biol.* **578**, 3–22 (2009).
19. Visscher, P. M. *et al.* 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
20. Guo, Q. *et al.* Identification of Novel Genetic Markers of Breast Cancer Survival. *J Natl Cancer Inst* **107**, (2015).
21. Morales, J. *et al.* A standardized framework for representation of ancestry data in genomics studies, with application to the NHGRI-EBI GWAS Catalog. *Genome Biol* **19**, 21 (2018).
22. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations: The American Journal of Human Genetics. [https://www.cell.com/ajhg/fulltext/S0002-9297\(17\)30107-6](https://www.cell.com/ajhg/fulltext/S0002-9297(17)30107-6).
23. Genome-wide association studies: implications for multiethnic samples - PMC. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782359/>.
24. Genetic Background of Patients from a University Medical Center in Manhattan: Implications for Personalized Medicine | PLOS ONE. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019166>.
25. Hc, S. *et al.* A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. *European journal of human genetics : EJHG* **24**, (2016).
26. Martin, A. R., Teferra, S., Möller, M., Hoal, E. G. & Daly, M. J. The critical needs and challenges for genetic architecture studies in Africa. *Current Opinion in Genetics & Development* **53**, 113–120 (2018).

27. International Schizophrenia Consortium *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
28. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
29. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
30. Talarico, F. [UNIFESP. O Uso Do Escore Poligênico Como Uma Ferramenta De Determinação De Risco De Esquizofrenia. (2018).
31. Leucht, S. *et al.* What does the PANSS mean? *Schizophrenia Research* **79**, 231–238 (2005).
32. Busner, J. & Targum, S. D. The Clinical Global Impressions Scale. *Psychiatry (Edgmont)* **4**, 28–37 (2007).
33. Pedersen, G., Urnes, Ø., Hummelen, B., Wilberg, T. & Kvarstein, E. H. Revised manual for the Global Assessment of Functioning scale. *Eur. Psychiatry* **51**, 16–19 (2018).
34. Desjardins, P. & Conklin, D. NanoDrop Microvolume Quantitation of Nucleic Acids. *J Vis Exp* (2010) doi:10.3791/2565.
35. Choi, S. W., Mak, T. S.-H. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* **15**, 2759–2772 (2020).
36. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, (2019).
37. Psychiatric Genomics Consortium. *Psychiatric Genomics Consortium* <https://www.med.unc.edu/pgc/>.

38. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgm.harvard.edu & Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
39. Walton, E. *et al.* Prefrontal inefficiency is associated with polygenic risk for schizophrenia. *Schizophr Bull* **40**, 1263–1271 (2014).
40. Whalley, H. C. *et al.* Polygenic risk and white matter integrity in individuals at high risk of mood disorder. *Biol Psychiatry* **74**, 280–286 (2013).
41. Whalley, H. C. *et al.* The influence of polygenic risk for bipolar disorder on neural activation assessed using fMRI. *Transl Psychiatry* **2**, e130 (2012).
42. Holmes, A. J. *et al.* Individual differences in amygdala-medial prefrontal anatomy link negative affect, impaired social functioning, and polygenic depression risk. *J Neurosci* **32**, 18087–18100 (2012).
43. Tomasik, J. *et al.* Association of Insulin Resistance With Schizophrenia Polygenic Risk Score and Response to Antipsychotic Treatment. *JAMA Psychiatry* **76**, 864–867 (2019).
44. Li, J., Yoshikawa, A., Brennan, M. D., Ramsey, T. L. & Meltzer, H. Y. Genetic predictors of antipsychotic response to lurasidone identified in a genome wide association study and by schizophrenia risk genes. *Schizophr. Res.* **192**, 194–204 (2018).
45. Cavalcante, D. *et al.* O8.2. DURATION OF UNTREATED PSYCHOSIS (DUP) IS ASSOCIATED WITH WORSE RESPONSE TO TREATMENT IN ANTIPSYCHOTIC NAÏVE FIRST EPISODE PSYCHOSIS. *Schizophr Bull* **44**, S96–S96 (2018).

46. Kraguljac, N. V. *et al.* White matter integrity, duration of untreated psychosis, and antipsychotic treatment response in medication-naïve first-episode psychosis patients. *Mol Psychiatry* **26**, 5347–5356 (2021).
47. Baselmans, B. M. L., Yengo, L., van Rheenen, W. & Wray, N. R. Risk in Relatives, Heritability, SNP-Based Heritability, and Genetic Correlations in Psychiatric Disorders: A Review. *Biol Psychiatry* **89**, 11–19 (2021).
48. Legge, S. E. *et al.* Genetic architecture of schizophrenia: a review of major advancements. *Psychological Medicine* **51**, 2168–2177 (2021).