

Avaliação quantitativa do sofrimento psíquico: curadoria e análise de dados de uma coorte de pacientes com transtorno bipolar do humor

Andrew B. Moreira¹, Júlio M. Almeida², Rodrigo Dias³, Cibele M. Russo¹

¹Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo (USP) - São Carlos, SP - Brasil

²Instituto de Matemática, Estatística e Ciência da Computação (IME)
Universidade de São Paulo (USP) - São Paulo, SP - Brasil

³Dynamic Mind Health - São Paulo, SP - Brasil

unkind_pear@usp.br, julioalmeida2004@usp.br,
rodrigo@dmhealth.com.br, cibeled@icmc.usp.br

Abstract. *The present study applies data science techniques to address challenges in the measurement of psychological suffering in longitudinal studies involving patients with multifactorial mental disorders and clinical data with missing values and inconsistencies. To this end, a hybrid curation pipeline was developed, combining longitudinal reconstruction and Multiple Imputation by Chained Equations (MICE). Using the clinical database of the Bipolar Disorder Program (PROMAN) at FMUSP, statistical analyses were conducted to investigate relationships among psychiatric measures. The results revealed coherent patterns involving comorbidities, mood episodes, hospitalizations, and history of suicide attempts.*

Resumo. *O presente estudo aplica técnicas de ciência de dados para lidar com problemas de mensuração do sofrimento psíquico em estudos longitudinais com pacientes portadores de transtornos mentais multifatoriais e dados clínicos com valores ausentes e inconsistências. Para isso, foi desenvolvido um pipeline híbrido de curadoria, combinando reconstrução longitudinal e Imputação Múltipla por Equações Encadeadas (MICE). Utilizando a base clínica do Programa de Transtorno Bipolar (PROMAN) da FMUSP, foram conduzidas análises estatísticas para investigar relações entre medidas psiquiátricas. Os resultados evidenciaram padrões coerentes entre comorbidades, episódios de humor, hospitalizações e histórico de tentativa de suicídio.*

1. Introdução

Este trabalho está inserido na temática da mensuração quantitativa do sofrimento psíquico, que confere por si só um desafio científico de grande complexidade. Diferentemente de variáveis biológicas como pressão arterial e glicemia, construtos psicológicos - como depressão, ansiedade e impulsividade - são, por natureza, abstratos, isto é, não são diretamente observáveis, de modo que dependem de variáveis latentes para a sua medição. Por isso, a difícil tarefa de tradução desses construtos em medidas quantitativas confiáveis

é objeto central de áreas inteiras, notadamente a psicometria e a avaliação psicológica [Gorenstein, Wang e Hungerbühler 2015].

Se isoladamente a mensuração do sofrimento já é um problema, do ponto de vista de saúde pública, a situação se agrava, pois a precisão diagnóstica determina se o paciente vai conseguir acessar o tratamento necessário para o seu caso, o que pode afetar o curso da doença. Nesse sentido, a qualidade do tratamento tem dependência direta na qualidade dos instrumentos de avaliação. No contexto de doenças psiquiátricas, o Transtorno Bipolar (TB) afeta cerca de 2,5% da população mundial [Bueno 2024], associado a perdas funcionais severas, alta taxa de hospitalização e risco aumentado de suicídio.

Logo, o aprimoramento da psicometria parece fundamental para a saúde mental pública. Contudo, quando lidamos em um cenário de dados clínicos longitudinais já coletados, com os instrumentos já determinados, não é mais possível aplicar um plano de avaliação psiquiátrica "perfeita" ou "ótima". Diversos problemas novos em dados do mundo real são identificados como empecilhos para a qualidade diagnóstica: protocolos extensos, preenchidos manualmente ao longo de anos por múltiplos profissionais, com inconsistências estruturais, além de valores ausentes e violações de regras de domínio.

Além dos problemas de qualidade dos dados, outro desafio importante consiste na relação entre as diferentes medidas psiquiátricas em transtornos mentais multidimensionais, como o TB, cuja dinâmica do humor é complexa e heterogênea, o que necessita o uso de diferentes instrumentos que capturem as distintas dimensões do transtorno. Assim, compreender as relações entre essas medidas pode fornecer subsídios relevantes para interpretação clínica e acompanhamento longitudinal dos sintomas.

Diante desse contexto, propomos um estudo voltado à aplicação de técnicas de ciência de dados para lidar com problemas de mensuração do sofrimento psíquico em bases clínicas longitudinais previamente coletadas. Além da curadoria e recuperação da base de dados, o trabalho também investigou relações entre medidas psiquiátricas por meio de análises estatísticas, sob a perspectiva de uma avaliação integrada do transtorno bipolar. Para alcançar tais objetivos e utilizando a base clínica longitudinal do Programa de Transtorno Bipolar (PROMAN) do Instituto de Psiquiatria do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, propomos o seguinte roteiro de tratamento e análise de dados:

- um *pipeline* reprodutível de curadoria de bases psiquiátricas longitudinais, combinando regras lógicas e imputação probabilística de dados via Imputação Múltipla por Equações Encadeadas (MICE) [van Buuren e Groothuis-Oudshoorn 2011];
- uma análise estatística de uma coorte de 139 pacientes com TB, focando na relação entre as medidas psiquiátricas a fim de identificar associações clinicamente interpretáveis;
- evidência empírica de que abordagens integradas de avaliação — consistentes com as recomendações de [Meyer et al. 2001] — são mais informativas que a busca por um instrumento isolado “ótimo”.

2. Fundamentação Teórica

A avaliação psicológica é um processo sistemático de coleta e integração de informações para compreender o funcionamento psicológico dos indivíduos, mediante instrumentos

padronizados e fontes complementares como entrevistas e observações clínicas [Meyer et al. 2001]. Para tanto, este trabalho se alinha com entendimento de que uma avaliação clinicamente útil emerge da integração de múltiplas fontes [Meyer et al. 2001], e não da escolha do instrumento perfeito. Esse princípio orienta a construção do protocolo do PROMAN, que combina escalas consagradas como a *Young Mania Rating Scale* (YMRS), a *Hamilton Depression Rating Scale* (HAM-D) e o *Global Assessment of Functioning* (GAF) [Gorenstein, Wang e Hungerbühler 2015].

Essa avaliação psíquica, ainda que integrada por vários processos, deve primordialmente ser baseada por instrumentos padronizados, dos quais a psicometria é responsável pela sua criação, construção e validação. Por isso, o arsenal psicométrico também é de importante fundamentação teórica para esse estudo. Parte desse arsenal é essencialmente composto pela definição clara do que está sendo medido, com a distinção da *definição constitutiva* e a *definição operacional*. Além disso, segundo Pasquali (2009), a elaboração de itens psicométricos deve seguir critérios de clareza, tipicidade e credibilidade, com avaliação por juízes especialistas para garantir sua adequação ao construto.

3. Trabalhos Relacionados

Para situar o estudo no estado da arte, adotou-se uma busca bibliográfica, nas bases PubMed, Scopus e Google Scholar, sistematizada por descritores, com correlatos e equivalentes em inglês: transtorno bipolar; Imputação Múltipla por Equações Encadeadas; curadoria de dados. Após isso, realizou-se uma seleção exploratória de artigos publicados a partir de 2010 e com escopo similar ao nosso estudo.

No domínio do tratamento de dados clínicos ausentes, van Buuren e Groothuis-Oudshoorn [2011] consolidam a Imputação Múltipla por Equações Encadeadas (MICE) como técnica de referência para bases multivariadas com padrões complexos de ausência. Ao contrário de imputações ingênuas (média, último valor observado), o MICE modela cada variável incompleta condicionalmente às demais, preservando estruturas de correlação. Sua implementação no `IterativeImputer` do `scikit-learn` permite aplicação reprodutível em bases psiquiátricas.

No contexto de aplicações de ciência de dados em Transtorno Bipolar (TB), Rothenberg et al. [2021] exploraram modelos de aprendizado de máquina (como *Random Forest* e SVM) para predição de recaídas depressivas em pacientes com transtorno bipolar, demonstrando a utilidade de algoritmos preditivos em dados clínicos longitudinais. O presente trabalho dialoga com a literatura ao propor um pipeline reprodutível de curadoria que combina a engenharia de dados à lógica clínica, servindo de etapa preparatória essencial antes de análises inferenciais ou preditivas.

4. Metodologia

4.1. Base de Dados e Auditoria Inicial

O núcleo empírico deste trabalho é uma base clínica longitudinal fornecida pelo PROMAN, contendo registros de 139 pacientes com TB avaliados por meio de protocolo de entrevista clínica estruturada. Todo o *pipeline* foi implementado em Python [Python Software Foundation 2026], utilizando `pandas`, `scipy`, `scikit-learn` e `matplotlib`.

A auditoria inicial revelou que 84 dos 139 pacientes (60,43%) apresentavam alguma inconsistência, totalizando 103 erros mapeados. A maioria (56,83%) estava ligada a descompassos na contagem de classes de medicações (`nclassemedicacao`). Identificaram-se também violações de domínios (`tip_subs`) e regras condicionais (uso de substância em contradição com a variável primária `subst_1`). As taxas de dados ausentes antes do tratamento eram críticas em diversas variáveis estruturais. A Tabela 1 detalha as variáveis essenciais para o estudo de associação, seu tipo, descrição e a proporção de incompletude original na base.

Tabela 1. Descrição das variáveis críticas analisadas e sua proporção de dados ausentes antes do tratamento ($N = 139$).

Variável	Tipo	Descrição	Ausentes (%)
<code>idade</code>	Quant. discreta	Idade atual do paciente	0,0%
<code>imc</code>	Quant. contínua	Índice de Massa Corporal (kg/m^2)	28,78%
<code>an_compl</code>	Quant. discreta	Anos completos de escolaridade	7,91%
<code>id_inic_doenca</code>	Quant. discreta	Idade de início dos sintomas de TB	10,79%
<code>hospit</code>	Qualit. binária	Histórico de hospitalização (Sim/Não)	35,97%
<code>quantas</code>	Quant. discreta	Número de hospitalizações psiquiátricas	73,38%
<code>id_1hos</code>	Quant. discreta	Idade da primeira hospitalização	69,78%
<code>tent_sui</code>	Qualit. binária	Histórico de tentativa de suicídio (Sim/Não)	38,13%
<code>n_tent</code>	Quant. discreta	Número de tentativas de suicídio	75,54%
<code>tept_at</code>	Qualit. binária	Presença atual de TEPT	95,68%

4.2. Pipeline de Curadoria

O *script* de limpeza foi estruturado em quatro fases consecutivas.

Fase 1 - Remoção de Variáveis Irrecuperáveis. Variáveis com 100% de ausência (`fer_cab`, `asma`, `diabetes`) foram descartadas por não conterem informação mínima útil.

Fase 2 - Imputação Conservadora por Regra. Para variáveis de contagem de episódios (`nepisodios`, `n_mania`, `n_depres`, `n_hipoma`, `n_mistos`), valores ausentes foram convertidos em zero. A decisão é justificada pela verificação aritmética: a soma dos subtotais de episódios coincide com `nepisodios` sob essa suposição. Igualmente, pacientes com `hospit=2` (sem hospitalização) e `tent_sui=2` (sem tentativas) tiveram seus campos cumulativos zerados. Embora clinicamente justificáveis para manter a coerência interna, essas imputações determinísticas podem introduzir vieses ao subestimar a real contagem caso episódios passados tenham sido esquecidos pelos pacientes.

Fase 3 - Reconstrução Longitudinal Lógica. O questionário separa sintomatologia atual e passada, mas o preenchimento gera inconsistências ao se reconstruir o histórico vitalício. Seja $X_{at} \in \{0, 1\}$ o indicador de presença atual e $X_{pas} \in \{0, 1\}$ o indicador no passado. Define-se a presença ao longo da vida (*lifetime*) como a disjunção lógica:

$$L(X) = X_{at} \vee X_{pas} = \max(X_{at}, X_{pas}). \quad (1)$$

Por essa lógica, foram reconstruídas 16 grandes síndromes psicopatológicas, incluindo Transtorno de Estresse Pós-Traumático (TEPT), Pânico (com e sem agorafobia), Transtorno de Ansiedade Generalizada (TAG), Fobia Social, Fobia Específica, Dependência e Abuso de Álcool/Substâncias, Transtorno Obsessivo-Compulsivo (TOC), Anorexia, Bulimia e Compulsão Alimentar Periódica. Sem essa reconstrução, exclusão ingênua por *pairwise deletion* eliminaria parte substancial da amostra por erros descritivos de preenchimento.

Fase 4 - Imputação Múltipla por Equações Encadeadas (MICE). Lacunas remanescentes foram tratadas com MICE [van Buuren e Groothuis-Oudshoorn 2011], via `IterativeImputer` do `scikit-learn`. Para um vetor $X = (X_1, \dots, X_p)$, a imputação de X_j com dados faltantes é modelada iterativamente pela distribuição condicional:

$$P(X_j \mid X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p, \theta_j). \quad (2)$$

O algoritmo foi parametrizado com 10 iterações cíclicas e semente aleatória fixa (42), garantindo reprodutibilidade. Como pós-processamento necessário, as variáveis qualitativas e discretas imputadas de forma contínua pelo MICE foram arredondadas para o inteiro mais próximo e limitadas (*clipped*) ao intervalo de valores permitidos no dicionário de variáveis para evitar valores inválidos (contagens negativas ou valores fracionários em variáveis binárias). Após o *pipeline* completo, atingiu-se 0% de dados faltantes nas variáveis matemáticas e clínicas estruturais, tornando a base apta para inferência.

4.3. Análise estatística

Inicialmente, foram conduzidas análises estatísticas univariadas para caracterizar descritivamente a amostra. Posteriormente, conduziu-se análises bivariadas. Nessa etapa foram excluídos pares de variáveis com dependência estrutural direta (famílias *lifetime*), a fim de evitar associações óbvias. Para os estudos de associação entre medidas psiquiátricas, empregaram-se testes inferenciais selecionados conforme a escala de mensuração das variáveis e justificados a seguir:

Teste Qui-Quadrado de Independência (χ^2): utilizado para investigar associações entre variáveis categóricas. A estatística é definida como $\chi^2 = \sum (O_i - E_i)^2 / E_i$. Esse teste é o padrão para tabelas de contingência, com as premissas de que pelo menos 80% das células esperadas tenham frequências ≥ 5 e nenhuma célula tenha frequência < 1 , garantidas em auditoria prévia.

Coefficiente de correlação de Spearman (ρ): adotado para avaliar relações monotônicas entre variáveis numéricas quantitativas, definido por $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$. Seu uso se justifica já que dados de contagem psiquiátrica violam fortemente a suposição de normalidade requerida para o coeficiente de Pearson.

Teste *t* de Welch: empregado para comparação de médias entre dois grupos, com escolha automática motivada por violação da homogeneidade de variâncias ou pela presença de amostras altamente desbalanceadas (como o grupo TB-I com $N = 127$ contra TB-II com $N = 12$), em que o teste *t* de Student convencional perde robustez. A normalidade das distribuições foi avaliada preliminarmente pelo teste de Shapiro-Wilk [Shapiro e Wilk 1965].

5. Resultados, Discussão e Limitações

5.1. Perfil geral da amostra

A amostra final ($N = 139$) apresentou predomínio feminino (66,9%) e étnico branco (79,9%), com equilíbrio entre casados e solteiros (41,0% cada). Do ponto de vista diagnóstico, 91,4% dos pacientes foram classificados como TB-I, sugerindo uma coorte de maior gravidade clínica, possivelmente relacionada ao perfil terciário de encaminhamento do PROMAN.

A idade média dos participantes foi de 40,05 anos ($\sigma = 11,35$; min=17, max=76), enquanto o início médio da doença ocorreu aos 22,42 anos ($\sigma = 10,18$), incluindo casos de início precoce. O IMC médio foi de 26,92 kg/m² ($\sigma \approx 4,48$), indicando tendência ao sobrepeso, achado clinicamente relevante diante da associação entre estabilizadores de humor, antipsicóticos atípicos e ganho de peso [Vancampfort et al. 2013]. A escolaridade média foi de 11,88 anos completos ($\sigma = 3,15$). O número de hospitalizações apresentou desvio padrão de $\sigma = 3,27$, enquanto a idade da primeira internação psiquiátrica variou entre 11 e 60 anos na base tratada.

Observou-se também impacto funcional importante: apenas 33,7% dos pacientes com ocupação informada relataram atividade remunerada ativa, enquanto os demais distribuíram-se entre auxílio-doença, aposentadoria por invalidez e atividades domésticas. Embora o estudo não tenha sido desenhado para mensurar funcionalidade ocupacional de maneira específica, esse padrão é compatível com a literatura sobre perdas funcionais associadas ao transtorno bipolar [Léda-Rêgo, Bezerra-Filho e Miranda-Scippa 2020].

5.2. Relações entre as medidas psiquiátricas

Esta subseção concentra as principais análises de associação entre medidas psiquiátricas, buscando fornecer subsídios para interpretação clínica e compreensão integrada do transtorno bipolar.

Inicialmente, investigou-se a consistência diagnóstica dos agrupamentos do TB, verificando sua conformidade com as definições psicopatológicas propostas pelo DSM-5 [American Psychiatric Association 2013]. Como mostra a Tabela 2, pacientes TB-I apresentaram média significativamente maior de hospitalizações que TB-II ($p < 0,001$), enquanto episódios de mania foram exclusivos do TB-I ($p < 0,001$) e episódios hipomaniacos concentraram-se no TB-II ($p = 0,018$).

Tabela 2. Comparação de médias entre subtipos diagnósticos (TB-I vs TB-II).

Marcador clínico	TB-I	TB-II	Valor- p
Quantidade de hospitalizações	2,08	0,33	$< 0,001$
Episódios de mania	4,44	0,00	$< 0,001$
Episódios de hipomania isolada	0,01	2,25	0,018

As análises de associação por Qui-Quadrado revelaram três agrupamentos clinicamente relevantes.

- (i) **Eixo Ansiedade-Comorbidade:** o transtorno de ansiedade ao longo da vida apresentou as associações mais expressivas com outras comorbidades psiquiátricas, incluindo TOC ($\chi^2 = 44,0$), Fobia Social ($\chi^2 = 33,1$), Fobia Específica ($\chi^2 = 31,4$)

e Pânico com Agorafobia ($\chi^2 = 28,1$), todos com $p < 0,001$. Os resultados são compatíveis com evidências de que sintomas ansiosos frequentemente coexistem na estrutura clínica do transtorno bipolar [Freeman et al. 2002].

- (ii) **Eixo Psicótico:** Transtorno de Humor Psicótico associou-se a manifestações psicóticas no primeiro episódio ($\chi^2 = 18,7$, $p < 0,001$), resultado compatível com modelos dimensionais e contínuos da psicopatologia afetiva e psicótica [Craddock et al. 2005].
- (iii) **Eixo de Vulnerabilidade Suicida:** tentativas de suicídio associaram-se a transtornos alimentares ao longo da vida ($\chi^2 = 9,6$, $p = 0,002$) e ao sexo ($\chi^2 = 9,6$, $p = 0,002$).

Por meio do coeficiente de correlação de Spearman, investigou-se a associação entre as duas principais manifestações do TB — mania e depressão — mensuradas pela contagem de episódios. Observou-se correlação positiva entre episódios maníacos e depressivos ($\rho = 0,44$, $p < 0,001$), resultado compatível com modelos de gravidade global e recorrência do curso do transtorno bipolar [Judd et al. 2002].

Além disso, idade atual correlacionou-se positivamente com número de hospitalizações ($\rho = 0,45$), enquanto idade de início da doença também apresentou associação com hospitalizações ($\rho = 0,32$).

Tabela 3. Correlações de Spearman significativas ($\rho \geq 0,30$, $p < 0,05$) na base tratada.

Variável 1	Variável 2	ρ	Valor- p
Idade atual	Número de hospitalizações	0,45	$< 0,001$
Número de Episódios de Mania	Número de Episódios Depressivos	0,44	$< 0,001$
Idade de início da doença	Número de hospitalizações	0,32	$< 0,001$

5.3. Precocidade clínica e vulnerabilidade suicida

De modo complementar, investigou-se o efeito da precocidade do transtorno bipolar sobre tentativas de suicídio, buscando identificar possíveis marcadores de maior vulnerabilidade clínica.

Para isso, empregou-se comparação de médias entre grupos definidos pelo histórico de tentativa de suicídio ao longo da vida. A normalidade das distribuições foi avaliada por meio do teste de Shapiro-Wilk [Shapiro e Wilk 1965]; diante de rejeição da hipótese de normalidade, optou-se pelo teste t de Welch.

Indivíduos com pelo menos uma tentativa apresentaram início dos sintomas significativamente mais precoce (17,22 anos vs. 25,72; $p < 0,001$). Esse grupo também acumulou maior número médio de comorbidades ao longo da vida (1,96 vs. 1,26; $p = 0,029$), iniciou medicação mais cedo (24,11 vs. 31,21; $p < 0,001$) e apresentou primeira hospitalização psiquiátrica em idade mais jovem (28,46 vs. 34,73; $p < 0,001$). Em conjunto, os resultados sugerem que precocidade clínica pode estar associada a trajetórias mais graves de adoecimento.

Tabela 4. Comparação de médias entre grupos (base: tentativa de suicídio ao longo da vida).

Marcador clínico	Tentaram	Nunca	Valor- <i>p</i>
Idade de início da doença (anos)	17,22	25,72	< 0,001
Idade início primeira medicação (anos)	24,11	31,21	< 0,001
Idade primeira hospitalização (anos)	28,46	34,73	< 0,001
Idade corrente (anos)	35,48	42,95	< 0,001
Idade de estabilização clínica (anos)	29,24	34,45	< 0,001
Contagem de comorbidades ao longo da vida	1,96	1,26	0,029

Nota: Valores-*p* bilaterais. Welch quando normalidade violada; Student caso contrário. A variável-desfecho refere-se à tentativa, não a óbito por suicídio. Não é possível inferir causalidade direta.

5.4. Avaliação do Impacto da Imputação e Análise de Sensibilidade

Para quantificar o efeito do pipeline de curadoria e responder a questionamentos metodológicos, realizou-se uma análise de sensibilidade comparando os resultados da base tratada por MICE com duas abordagens alternativas: a exclusão pareada (*pairwise deletion*) nos dados brutos originais e a imputação simples pela média. A Tabela 5 resume o impacto dessas estratégias sobre os principais coeficientes de correlação de Spearman (ρ).

Tabela 5. Comparação dos coeficientes de correlação de Spearman (ρ) sob diferentes métodos de tratamento de dados ausentes.

Pares de Variáveis	Dados Brutos (Pareada)	Imputação pela Média	Pipeline MICE (Proposto)
Idade atual $\times$ Hospit.	0,146 ($p = 0,389$)	-0,017 ($p = 0,840$)	0,448 ($p < 0,001$)
Ep. Mania $\times$ Ep. Depres.	0,491 ($p < 0,001$)	0,336 ($p < 0,001$)	0,442 ($p < 0,001$)
Início TB $\times$ Hospit.	0,170 ($p = 0,313$)	-0,010 ($p = 0,907$)	0,321 ($p < 0,001$)

Os resultados evidenciam que a exclusão pareada reduz drasticamente o tamanho amostral efetivo (para $N = 37$ nos cruzamentos que envolvem hospitalizações), o que anula o poder estatístico e torna as associações não-significativas. Por outro lado, a imputação simples pela média artificialmente colapsa a variabilidade dos dados e dilui as correlações para próximo de zero. O pipeline MICE proposto, ao estimar os valores faltantes condicionalmente ao perfil clínico multivariado do paciente, preservou a estrutura de associação dos dados e recuperou o sinal clínico em toda a amostra ($N = 139$).

Essa diferença de sensibilidade também é nítida nos testes inferenciais de médias. No teste Welch T-test para comparar a idade de início dos sintomas de TB entre pacientes com e sem tentativa de suicídio, a exclusão pareada nos dados brutos ($N = 86$ válidos) indicou diferença significativa (18,24 vs. 23,42 anos; $p = 0,006$). Com a aplicação do MICE ($N = 139$), a diferença de médias acentuou-se (17,22 vs. 25,72 anos), e a significância estatística aumentou para $p = 7,6 \times 10^{-8}$, demonstrando que a curadoria reconstrói relações clínicas consistentes que estavam obscurecidas pelas omissões de registro.

5.5. Discussão de Limitações e Vieses

Apesar dos resultados clinicamente coerentes e estatisticamente robustos, o estudo possui limitações importantes. A amostra é relativamente pequena ($N = 139$) e proveniente de um único centro especializado terciário (PROMAN no HC-FMUSP). Isso pode introduzir viés de seleção (pacientes com maior gravidade e comorbidades do que a população geral) e limitar a generalização dos achados para a atenção primária de saúde mental.

Além disso, parte da estrutura dos dados foi reconstruída por regras lógicas e técnicas de imputação. Embora a utilização de MICE permita preservar relações multivariadas entre variáveis clínicas, o método pressupõe que os mecanismos de ausência sejam ao menos parcialmente explicáveis pelas variáveis observadas (mecanismo MAR). Em bases psiquiátricas, a ausência de dados pode refletir fatores clínicos ou institucionais não observados, potencialmente introduzindo vieses residuais. Adicionalmente, as imputações determinísticas da Fase 2 (como assumir zero episódios para campos ausentes) são conservadoras, mas podem subestimar a real gravidade clínica se houver viés de recordação por parte dos pacientes. Assim, os resultados devem ser interpretados como evidências exploratórias e geradoras de hipótese, e não como conclusões causais definitivas.

6. Conclusão e Trabalhos Futuros

Este trabalho alcançou plenamente o seu objetivo geral de investigar quantitativa e psicometricamente a mensuração do sofrimento psíquico no transtorno bipolar, integrando a curadoria de dados ruidosos e análises estatísticas inferenciais.

Especificamente, os objetivos específicos foram atendidos da seguinte forma: (i) propôs-se e validou-se um pipeline reproduzível de curadoria que recuperou 103 inconsistências e imputou dados ausentes de variáveis com até 95% de incompletude sem distorcer o sinal clínico original; (ii) realizou-se a análise estatística inferencial de uma coorte de 139 pacientes do PROMAN, evidenciando trajetórias clínicas consistentes; e (iii) obteve-se evidência empírica corroborando que a avaliação baseada em múltiplos instrumentos integrados é estatisticamente mais informativa do que a busca por uma única escala isolada. Para fomentar a reprodutibilidade em pesquisas de saúde mental, o pipeline completo e os scripts de tratamento de dados foram disponibilizados publicamente em <https://github.com/unkind-pear/pipeline-dados>.

Como trabalhos futuros decorrentes deste estudo, planeja-se: (a) treinar modelos preditivos de aprendizado de máquina (*Random Forest* e *gradient boosting*) na base tratada para prever a gravidade de novos episódios e o risco de internação; (b) aplicar análise de sobrevivência (modelos de Kaplan-Meier e regressão de Cox) para modelar longitudinalmente o tempo até a recaída clínica; e (c) empregar a Teoria de Resposta ao Item (TRI) para avaliar as propriedades psicométricas individuais dos itens que compõem as escalas YMRS, HAM-D e GAF, refinando a qualidade do protocolo de coleta.

Reconhecimentos

O primeiro autor agradece ao CNPq pela bolsa do Programa PIBIC. O segundo autor agradece o Programa Unificado de Bolsas da USP. Os autores reconhecem a contribuição do Programa de Transtornos do Humor pela autorização do uso da base de dados. A ferramenta de IA generativa ChatGPT (OpenAI) foi utilizada para apoio no refinamento

linguístico e na edição textual do manuscrito. Os autores assumem total responsabilidade pelo conteúdo do trabalho, incluindo todas as análises, interpretações e conclusões.

Referências

- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. 5th ed. Washington, DC: American Psychiatric Publishing.
- Bueno, F. (2024). Transtorno bipolar afeta 140 milhões de pessoas no mundo e tem difícil diagnóstico. *Jornal da USP*, São Paulo, 14 mar. 2024.
- Bussab, W. O., & Morettin, P. A. (2017). *Estatística Básica*. 9. ed. São Paulo: Saraiva.
- Craddock, N., O'Donovan, M. C., & Owen, M. J. (2005). The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *Journal of Medical Genetics*, 42(3), 193–204.
- Freeman, M. P., Freeman, S. A., & McElroy, S. L. (2002). The comorbidity of bipolar and anxiety disorders: prevalence, psychobiology, and treatment issues. *Journal of Affective Disorders*, 68(1), 1–23.
- Gorenstein, C., Wang, Y. P., & Hungerbühler, I. (2015). *Instrumentos de Avaliação em Saúde Mental*. Porto Alegre: Artmed.
- Judd, L. L., Akiskal, H. S., Schettler, P. J., Endicott, J., Maser, J., Solomon, D. A., Leon, A. C., Rice, J. A., & Keller, M. B. (2002). The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Archives of General Psychiatry*, 59(6), 530–537.
- Léda-Rêgo, G., Bezerra-Filho, S., & Miranda-Scippa, Â. (2020). Functioning in euthymic patients with bipolar disorder: A systematic review and meta-analysis using the Functioning Assessment Short Test. *Bipolar Disorders*, 22(6), 569–581. <https://doi.org/10.1111/bdi.12904>
- Meyer, G. J. *et al.* (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56(2), 128–165.
- Pasquali, L. (2009). *Instrumentação Psicológica: Fundamentos e Práticas*. Porto Alegre: Artmed.
- Python Software Foundation (2026). *Python Language Reference*. Disponível em: <https://www.python.org>.
- Rotenberg, L. S., Borges-Júnior, R. G., Lafer, B., Salvini, R., & Dias, R. S. (2021). Exploring machine learning to predict depressive relapses of bipolar disorder patients. *Journal of Affective Disorders*, 295, 681–687. <https://doi.org/10.1016/j.jad.2021.08.127>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611.
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67.
- Vancampfort, D., Vansteelandt, K., Correll, C. U., Mitchell, A. J., De Herdt, A., Sienaert, P., Probst, M., & De Hert, M. (2013). Metabolic syndrome and metabolic abnormalities in bipolar disorder: a meta-analysis of prevalence rates and moderators. *American Journal of Psychiatry*, 170(3), 265–274.