

Hippocampal and Prefrontal MRI Alterations as Candidate Severity Biomarkers in Schizophrenia and Treatment Resistant Depression: A critical review with structured synthesis of published meta analyses

DOI: 10.5281/zenodo.23550269

Recebido: 19 out 2026

Aprovado: 20 out 2026

Danilo Ferreira de Sousa

Postdoctoral training at Universidade Federal de Alfenas.
PhD in Nursing, Universidade Federal do Ceará. Email:
prof.danilofersousa@gmail.com. ORCID: 0000-0001-9281-
165X.

Maria Isabel Pires Teixeira

Medical student, Faculdade de Petrolina, Petrolina,
Pernambuco, Brazil. Email: belzinhapt@gmail.com.

Regina Petrola Bastos Rocha

PhD in Health Sciences, Faculdade de Medicina do ABC.
Email: rpetrola7@gmail.com. ORCID: 0000-0003-0626-
232X.

Rafaela Nonato de Menezes

Postgraduate Program in Biotechnology, Universidade
Federal do Ceará. Email: rafaela.nonato@hotmail.com.
ORCID: 0000-0001-7468-3391.

ABSTRACT

Hippocampal and prefrontal abnormalities are repeatedly observed in psychiatric neuroimaging, but their value as individual markers of severity remains uncertain. This review synthesizes PubMed-indexed structural and functional MRI evidence in schizophrenia and treatment-resistant depression, with bibliographic verification on 7 October 2026. Published meta-analyses, large harmonized imaging studies, and selected treatment-response reports were examined through a framework separating case-control differences, symptom associations, functional outcomes, and prospective prediction. No new cross-disorder pooled effect was calculated. Large schizophrenia analyses reported reduced hippocampal volume and cortical thickness, whereas major-depression studies generally reported smaller structural effects. These major-depression findings were treated as contextual evidence rather than as treatment-resistant-depression estimates. Direct treatment-resistance reviews identified hippocampal and network alterations but showed substantial methodological variability; an exploratory coordinate-based meta-analysis produced limited modality-specific convergence. Functional connectivity and prefrontal treatment-target studies suggest clinically relevant circuits, yet response prediction differs from severity measurement. Medication exposure, illness duration, motion, clinical heterogeneity, and overlapping cohorts constrain interpretation. Hippocampal and prefrontal measures are therefore credible research candidates, but a reproducible group association does not establish a calibrated individual severity test. Clinical translation requires explicit target outcomes, longitudinal reliability, independent validation, comparison with clinical assessment, and demonstrated decision benefit.

Keywords: Magnetic resonance imaging, Hippocampus, Prefrontal cortex, Schizophrenia, Treatment resistant depression.

1 INTRODUCTION

Psychiatric disorders are diagnosed and monitored primarily through clinical history, observed behavior, patient-reported experience, and functional assessment. MRI offers another level of observation by measuring brain structure and aspects of physiological organization. The appeal of an imaging severity marker is clear: it might improve prognosis, identify treatment-relevant mechanisms, or provide an objective complement to symptoms. However, biological plausibility is not sufficient for clinical utility.

The hippocampus and prefrontal cortex are prominent candidates because they participate in memory, contextual processing, executive control, and emotional regulation. These functions are relevant to schizophrenia and persistent depressive illness, but they do not map onto a single clinical severity scale. A structural difference between diagnostic groups may reflect accumulated illness burden or other exposures rather than current symptom intensity.

Large collaborative studies have improved confidence that some group-level structural differences are reproducible. They have also shown that effects can be small, distributed, and influenced by demographic or clinical variables. This combination is scientifically valuable while remaining insufficient to define an individual diagnostic or severity threshold (van Erp et al., 2016; Schmaal et al., 2016).

Treatment-resistant depression presents an additional challenge. A history of inadequate response to appropriate treatment is a clinical course characteristic, not simply a high depression score. Evidence from general major depressive disorder cannot automatically be transferred to a population defined by treatment resistance. Similarly, a marker that predicts response to a particular intervention need not measure overall illness severity.

The objectives of this review are to summarize published structural and functional evidence involving hippocampal and prefrontal systems, distinguish clinically different biomarker claims, and assess the requirements for translation. The synthesis is restricted to verified PubMed-indexed publications and explicitly separates direct treatment-resistant-depression evidence from contextual major-depression findings.

2 THEORETICAL FRAMEWORK

2.1 Structure function and circuit interpretation

Structural MRI measurements include regional volume, cortical thickness, surface area, shape, and subfield estimates. They describe related but nonidentical properties. A reduced volume can arise through

changes in several components, while thickness and surface area have distinct developmental and measurement relationships. Combining them into one generic structural deficit loses information about what the study actually measured.

Functional MRI commonly assesses task-related blood-oxygenation signals or statistical relationships between regional time series at rest. Functional connectivity is a measure of covariation under a particular acquisition and processing pipeline. It does not directly establish anatomical communication, direction of influence, or a causal mechanism. Apparent increases and decreases can also depend on the reference comparison and preprocessing decisions.

The hippocampus is heterogeneous. Whole-structure measurements can conceal subfield patterns, while automated subfield labels require adequate image quality and consistent segmentation. Sun et al. (2023) reported both shared and divergent hippocampal subfield findings in schizophrenia and major depression. These results support anatomical specificity but do not establish a direct comparison between schizophrenia and treatment-resistant depression.

The prefrontal cortex is similarly heterogeneous. Dorsolateral, ventrolateral, orbitofrontal, and medial regions contribute differently to cognition and affective processing. A treatment target in dorsolateral prefrontal cortex can be clinically relevant through its connectivity with another region rather than through a local structural abnormality alone (Fox et al., 2012).

2.2 From an association to a biomarker

A biomarker claim must identify its intended use. Group discrimination, current symptom assessment, future outcome prediction, and treatment selection are different targets. Each requires an appropriate study design and validation strategy. A reproducible association is a foundation for this work, not its endpoint.

2.3 Defining severity and treatment resistance

Severity can refer to current symptoms, cognitive impairment, disability, relapse burden, chronicity, or need for care. These dimensions overlap but are not interchangeable. In schizophrenia, positive and negative symptoms can have different relationships with cognition and functional outcome. In depression, symptom intensity, recurrence, and treatment resistance can also diverge.

A structural measure linked to illness duration might index accumulated burden rather than present severity. A functional measure associated with treatment response might identify a modifiable circuit rather

than a general disease-state marker. Interpretation should begin with the clinical outcome being measured, not with the brain region attracting interest.

Table 1. Distinct clinical claims for an MRI measure

Claim	Necessary comparison	Main validation requirement
Case-control correlate	Patients versus appropriate controls	Replication and characterization of overlap
Current severity marker	Imaging versus prespecified clinical severity	Reliable association and individual calibration
Functional outcome correlate	Imaging versus real-world functioning	Specificity beyond symptoms and cognition
Prognostic marker	Baseline imaging versus future course	Prospective validation and calibration
Treatment-response predictor	Baseline imaging versus response to a defined treatment	Independent validation and treatment specificity
Treatment-selection marker	Differential benefit across interventions	Evidence of an interaction with treatment choice

Source: Review interpretation framework applied to the cited MRI literature.

Treatment-resistant depression should be identified through documented treatment adequacy and nonresponse, rather than inferred from chronic symptoms alone. The selected reviews use different study definitions, which limits comparability. The manuscript therefore preserves each source's clinical framing and treats historical chronic-depression findings as context where contemporary resistance criteria are not fully established.

3 METHODOLOGY

3.1 Design and retrieval

This manuscript is a structured critical review with synthesis of published meta-analytic findings. It does not report a registered systematic review, an original coordinate-based meta-analysis, or a new pooled correlation between MRI and severity. PubMed records and published summaries were verified on 7 October 2026.

The principal query combined schizophrenia or treatment-resistant depression with hippocampal or prefrontal terms and MRI, magnetic resonance, or meta-analysis terms in title and abstract fields. A

publication-date limit ended on the verification date. The principal query returned 2,933 records; twenty publications were retained for this critical synthesis. The broad search was supplemented by targeted retrieval of major collaborative structural studies, functional meta-analyses, and treatment-response publications. The query yield is reported separately from the selected reference set because exhaustive screening was not performed.

Priority was given to direct MRI evidence in the two clinical populations, published meta-analyses, and large harmonized multicenter analyses. General major-depression studies were included only as contextual comparisons and were labeled accordingly. Selected historical and treatment-target studies were used to clarify course and prediction claims. Nonindexed material, animal research, and unsupported clinical thresholds were not used.

3.2 Extraction and appraisal

The extraction basis was verified PubMed information and available published summaries. Recorded fields included diagnosis, design, imaging measure, comparator, clinical target, participant counts, and reported numerical estimates. Where confidence intervals for a particular structural effect were unavailable in the verified summary, no interval was invented or reconstructed from nominal sample counts.

Appraisal emphasized clinical applicability, outcome definition, confounding, multiplicity, cohort overlap, and validation. Formal risk-of-bias scoring and certainty grading were not conducted. This review consequently provides an organized interpretation of source findings rather than a validated evidence-quality ranking.

3.3 Quantitative synthesis rules

The selected evidence uses several incompatible statistical structures. Standardized case-control differences, symptom correlations, coordinate convergence, and connectivity-based treatment predictions answer different questions. They were not combined into a cross-disorder pooled effect. Such a calculation would produce a precise-looking number without a coherent clinical meaning.

The structural plot displays reported standardized point estimates and explicitly identifies general major-depression populations. It omits uncertainty intervals when those intervals were not available in the verified summary. The functional synthesis is organized through comparison and clinical target rather than through a numerical ranking of heterogeneous activation findings.

Table 2. Included evidence streams and permitted interpretation

Evidence stream	Principal example	Interpretive role
Schizophrenia structure	ENIGMA subcortical and cortical analyses	Robust group-level morphology
General depression structure	ENIGMA major-depression analyses	Contextual comparison, not a TRD effect
Direct TRD synthesis	Miola et al. 2023 and Carter et al. 2026	Resistance-related structural and functional evidence
Functional outcome	Wojtalik et al. 2017	Clinical functioning associations
Executive task activation	Minzenberg et al. 2009	Cognitive-control network dysfunction
Treatment-target connectivity	Fox et al. 2012 and Weigand et al. 2018	Prediction and targeting, not global severity

Source: The listed PubMed-indexed publications. TRD, treatment-resistant depression.

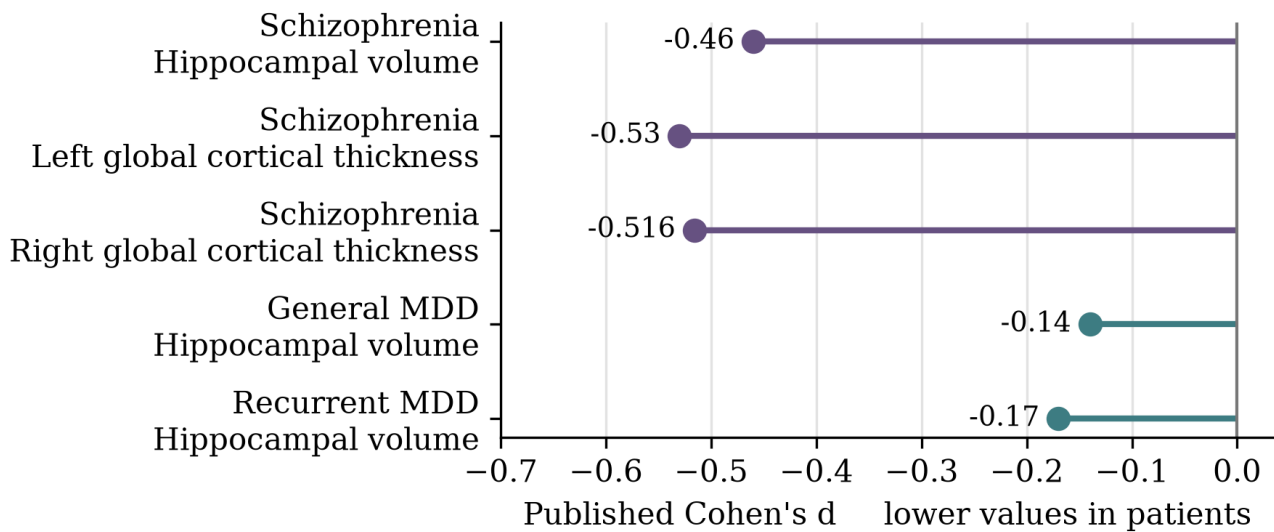
Published syntheses can share participants or primary reports, particularly within collaborative imaging consortia. Their sample sizes therefore were not summed to create a total unique population. Agreement across publications is interpreted as consistency of findings within potentially overlapping evidence rather than as guaranteed independent replication.

4 RESULTS AND DISCUSSION

4.1 Hippocampal and cortical structure in schizophrenia

The ENIGMA subcortical analysis by van Erp et al. (2016) reported smaller hippocampal volumes in schizophrenia, with a standardized effect of -0.46. This is evidence of a group difference after the source analysis's covariate framework; it is not a severity correlation. The accompanying associations with clinical and medication characteristics also emphasize that the signal arises within a heterogeneous illness population.

Figure 1. Selected published structural point estimates by diagnosis and measure.



Source: van Erp et al. (2016, 2018) and Schmaal et al. (2016). Values are reported Cohen's d estimates. Cortical entries are global hemisphere measures, not isolated prefrontal effects. Depression entries concern general MDD, not TRD. Confidence intervals are not supplied in this display.

The cortical analysis by van Erp et al. (2018) found widespread thinning, including prominent frontal and temporal effects. Associations with symptom severity were reported for regional thickness, but the published group effects cannot be converted into calibrated individual symptom estimates. Global hemisphere effects are shown only to convey the scale of the structural findings, not to represent a prefrontal biomarker threshold.

4.2 General depression findings provide context

Schmaal et al. (2016) reported a smaller hippocampus in major depressive disorder, with an overall standardized effect of -0.14. Recurrent and earlier-onset groups showed stronger reported differences than the overall group. Importantly, current symptom severity was not associated with regional volumes in that analysis. This directly limits a simple interpretation of hippocampal size as a contemporaneous depression-severity marker.

The cortical study by Schmaal et al. (2017) reported small adult differences involving orbitofrontal and other regions, with developmental variation across adult and adolescent analyses. These findings illustrate that age and illness stage can change the anatomical pattern. They should not be used as a treatment-resistant-depression estimate without direct evidence from that subgroup.

Ho et al. (2022) examined subcortical shape and localized differences to parts of the hippocampal and amygdala surfaces in clinically defined subgroups. Shape analysis increases anatomical detail, but a more specific anatomical result is not automatically a more useful individual predictor. Reliability and clinical calibration remain necessary.

Negative or weak findings also matter. De Kovel et al. (2019) did not find robust structural asymmetry alterations in a large major-depression analysis. Han et al. (2021) reported a small brain-age difference with substantial group overlap and no clear dependence on the examined clinical severity characteristics. These reports discourage interpreting every plausible imaging feature as a usable clinical marker.

Together, the general-depression literature supports subtle structural associations and clinically relevant heterogeneity. It does not establish that a particular hippocampal or prefrontal value grades the severity of refractory depression. Direct resistance-defined comparisons and prospective validation are required for that claim.

4.3 Direct evidence in treatment resistant depression

The historical study by Shah et al. (1998) compared chronic depressed patients with recovered patients and healthy controls and reported reduced left temporal gray-matter density including the hippocampus. It contributed an early anatomical hypothesis, but its small sample and historical chronicity framing restrict transfer to current resistance-defined populations.

Miola et al. (2023) reviewed structural and resting-state studies directly addressing treatment-resistant depression. Only eight studies met their criteria, and modality-specific coordinate-based analyses did not yield statistically significant convergence. An all-effects analysis implicated a precentral/superior frontal region. This supports cautious circuit-level hypotheses rather than a definitive hippocampal or prefrontal severity test.

Table 3. Direct TRD syntheses and their implications

Study	Evidence structure	Main interpretive implication
Klok et al. 2019	Review of structural MRI studies	Some abnormalities overlap with milder depression; specificity remains unresolved
Miola et al. 2023	Systematic review and exploratory coordinate-based meta-analysis	Limited convergence and a small heterogeneous evidence base

Study	Evidence structure	Main interpretive implication
Carter et al. 2026	Review of MRI comparisons with other depressive subgroups	Hippocampal reduction and altered networks remain methodologically variable
Yun et al. 2024	Review of MRI correlates of ketamine response	Treatment-linked candidates require distinction from severity markers

Source: Verified PubMed records of the listed studies.

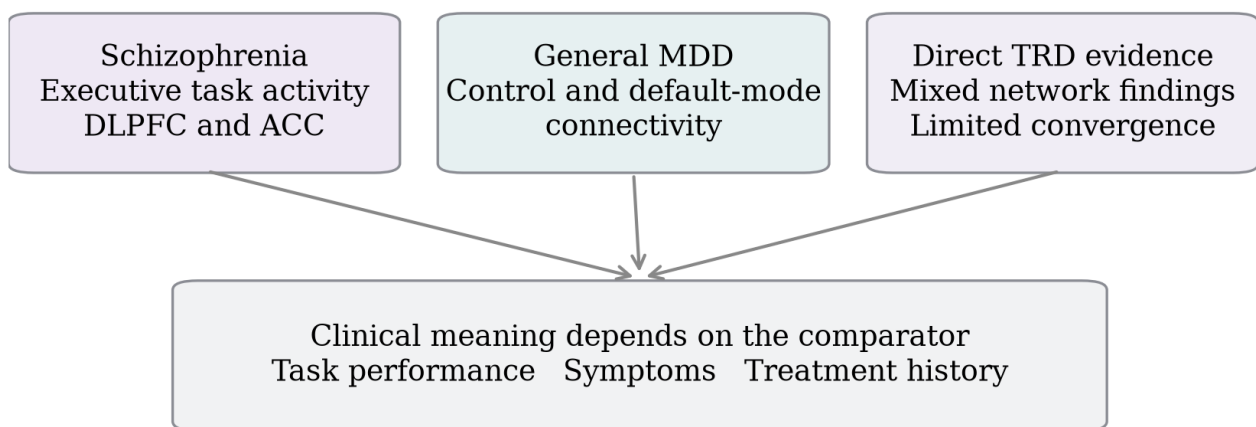
Carter et al. (2026) identified reduced hippocampal volume as a relatively consistent direct resistance-related finding, while describing mixed functional-connectivity patterns. The broader evidence base improves coverage but does not eliminate variation in comparator definitions, treatment history, and measurement. The reviews should be read as evolving interpretations of heterogeneous studies, not as independent confirmation of a validated clinical test.

4.4 Prefrontal functional organization

In schizophrenia, Minzenberg et al. (2009) synthesized executive-task imaging studies and identified altered engagement of a cognitive-control network, including dorsolateral prefrontal and anterior cingulate regions. Reduced activation in some nodes coexisted with increases elsewhere. This pattern is more informative than a uniform label of low frontal activity, because task demands and compensatory recruitment can influence the observed contrast.

In general major depression, Kaiser et al. (2015) reported imbalanced resting-state relationships among control, default-mode, and other networks. These findings support a distributed account of emotional and cognitive dysfunction. They do not establish a single sign of connectivity change that can be applied universally to an individual with refractory depression.

Figure 2. Circuit findings organized by disorder and clinical comparison.



Source: Qualitative schematic based on Minzenberg et al. (2009), Kaiser et al. (2015), Miola et al. (2023), and Carter et al. (2026). Boxes summarize evidence categories; arrows organize relationships and do not encode causal direction or measured connection strength.

Direct TRD syntheses report both increased and decreased connectivity depending on the comparison and network. Divergent signs do not necessarily represent contradictory biology: the same broad system can show different relationships across nodes, states, and analytic definitions. Nevertheless, this variability complicates translation into a stable severity scale.

The relevant next step is to connect a reproducible functional measure to an explicit clinical target, with reliability and validation established before a threshold is proposed.

4.5 Clinical functioning and symptom specificity

Wojtalik et al. (2017) synthesized brain relationships with functional outcome in schizophrenia and associated better functioning with greater frontolimbic and whole-brain volumes, smaller ventricles, and patterns of activation. This literature addresses a clinically meaningful dimension often missed by diagnosis-focused analyses. However, functioning can depend on social opportunity, cognition, symptoms, and support in addition to neurobiology.

An imaging association with real-world functioning therefore should not be interpreted as a direct reading of disease severity. The functional measure, follow-up interval, and environmental context need to be specified. A marker could add useful prognostic information while having a weak relationship with current symptoms, or vice versa.

Leerssen et al. (2020) illustrate symptom specificity in major depression. Associations with insomnia severity involved selected cortical surface-area measures and were not reproduced for overall depression severity. The result argues for studying explicit symptom dimensions rather than assuming that a global scale is the biological target of every imaging feature.

Table 4. Clinical endpoints and common inference errors

Endpoint	Appropriate interpretation	Inference to avoid
Diagnostic status	Group-associated neurobiology	An individual diagnostic threshold
Current symptom score	A state-related association	A validated prediction without calibration
Illness duration	Possible accumulated burden	Current severity measured directly
Functional outcome	Disability-related or prognostic relevance	Purely biological determination of functioning
Insomnia severity	Symptom-specific association	A marker of all depressive symptoms
Treatment response	Intervention-specific prediction	A universal severity index

Source: Interpretation informed by Wojtalik et al. (2017), Schmaal et al. (2016), and Leerssen et al. (2020).

Explicit endpoint definitions are therefore central to biomarker development. They determine whether a study contributes to severity monitoring, prognosis, or treatment selection.

4.6 Treatment prediction and prefrontal targets

Fox et al. (2012) related reported antidepressant efficacy of dorsolateral prefrontal stimulation sites to their functional relationship with the subgenual cingulate. The study supported a connectivity-based targeting hypothesis, initially using healthy-subject imaging and a smaller depressed comparison sample. This is a circuit and treatment-target claim rather than a general severity claim.

Weigand et al. (2018) prospectively examined whether target connectivity predicted antidepressant response. Their results strengthened the clinical relevance of the circuit, while the cohort sizes and study context remained important. Predicting response within a specified treatment setting does not establish that the same measure predicts natural course, relapse, or severity across all treatments.

Yun et al. (2024) reviewed MRI-based correlates of ketamine response in treatment-resistant depression. Candidate features involved several networks and symptom dimensions, including limbic and

frontoparietal systems. The diversity of features suggests that different symptom improvements and interventions may involve different aspects of the same broad circuitry.

This literature provides a more specific route to clinical utility than simply seeking the largest case-control abnormality. A marker may become useful when it identifies a modifiable treatment-relevant network. Even then, its incremental benefit must be established relative to existing clinical information and alternative targeting strategies.

A treatment-selection biomarker requires evidence that outcomes differ according to both marker status and the chosen intervention. A response association within one treated cohort is not sufficient to establish that interaction. Prospective comparative designs are needed to show whether using the marker actually improves patient outcomes, reduces ineffective treatment, or supports a better treatment choice.

The distinction between target relevance, response prediction, and severity measurement should remain explicit throughout translation. These roles may eventually converge in a clinically useful model, but their equivalence should be demonstrated rather than assumed.

4.7 Confounding and measurement reliability

Medication exposure is a major concern in structural psychiatric MRI. The cortical schizophrenia analysis reported larger case-control effects in medicated than unmedicated groups and associations with medication dose. These observations do not distinguish medication effects from illness severity or treatment selection, because those factors are interrelated (van Erp et al., 2018).

Illness duration, age at onset, recurrence, and comorbidity can similarly influence morphology. Childhood adversity and other exposures can contribute to relevant structural variation without being specific to the diagnosis or resistance status. Frodl et al. (2017) examined adversity-related effects within a large major-depression collaboration, illustrating the need to model exposures beyond diagnostic categories.

Functional MRI adds sensitivity to motion, vigilance, acquisition duration, physiological signals, and preprocessing choices. A correlation with symptoms can partly reflect a systematic difference in movement or state during scanning. Removing obvious artifacts is necessary but does not establish that a feature is stable enough for individual monitoring.

Table 5. Sources of variation requiring explicit handling

Source	Possible impact	Recommended reporting
Medication and treatment history	Structure and functional state	Exposure, timing, dose framework, and sensitivity analyses
Illness course	Accumulated burden and subgroup differences	Duration, recurrence, onset, and resistance definition
Head motion and vigilance	Functional signal distortion	Acquisition quality, exclusion, and residual associations
Scanner and segmentation	Measurement comparability	Harmonized acquisition and processing details
Clinical heterogeneity	Different targets within a diagnosis	Prespecified symptom and functional endpoints
Adversity and comorbidity	Nonspecific structural variation	Relevant measured exposures and adjustment rationale

Source: Critical interpretation of the selected structural, functional, and clinical imaging publications.

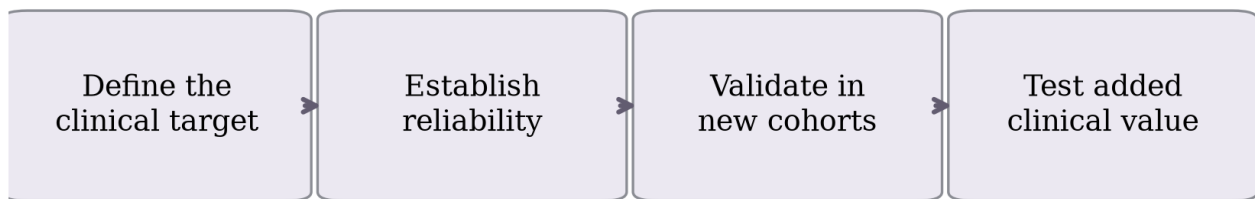
Reliability is especially important for a proposed monitoring biomarker. A feature that changes with scanner conditions more than with clinical course cannot reliably grade severity, even if its group association is statistically significant.

4.8 Group overlap and individual prediction

Large sample sizes can detect small differences that have limited value for classifying or monitoring an individual. A standardized mean difference quantifies separation between group averages relative to variability. It does not provide a person's probability of severe illness, a decision threshold, or the expected error of an individualized prediction.

The same issue applies to anatomical specificity. A hippocampal subfield finding may sharpen a mechanistic hypothesis while remaining too overlapping or unstable for clinical use. Conversely, a multivariable model can yield useful prediction without a large effect in any single region. The development question should therefore concern reproducible performance against an explicit target, not just the prominence of a named structure.

Figure 3. Validation pathway for a candidate MRI severity biomarker.



A reproducible group effect alone does not establish an individual severity test

Source: Conceptual methodological synthesis of the cited evidence. The diagram is a proposed evaluation pathway, not an observed trial result or an established clinical guideline.

Validation should examine discrimination, calibration, reliability, and incremental value. A model can rank patients correctly while systematically overestimating severity. It can also perform well within one cohort and fail in another with different scanners, clinical distributions, or treatment histories. Both failures matter when the intended use is individual decision-making.

The clinical reference assessment should be prespecified and obtained independently where feasible. If imaging is intended to improve prognosis, future outcomes should be measured after the scan. If it is intended to monitor severity, repeated imaging and clinical assessments are needed to establish meaningful within-person change.

Finally, clinical usefulness requires a plausible action linked to the prediction. A technically accurate marker may add little if it does not improve decisions or outcomes beyond a careful clinical assessment.

4.9 Research priorities

The field would benefit from studies that declare the intended biomarker role before analysis. The hippocampal and prefrontal focus should be justified by a clinical target and tested against appropriate alternative or whole-brain models. Selecting the best-performing region after exploring many possibilities can overstate specificity and predictive accuracy.

Table 6. Requirements for clinically credible validation

Requirement	Design feature	Clinical value established
Clear target	Prespecified symptom, disability, or prospective endpoint	Interpretability of the marker
Direct TRD applicability	Documented adequate treatment failures	Evidence in the intended population
Reliable measurement	Repeated acquisition or reproducibility assessment	Meaningful individual change
Independent testing	Held-out sites and external cohorts	Transportability
Clinical comparison	Model with and without existing clinical variables	Incremental information
Decision evaluation	Prospective testing of marker-informed care	Improved patient outcomes

Source: Research implications derived from the selected imaging evidence.

Prospective multisite studies should include sufficient variation in illness course and treatment exposure to test generalizability. Models should avoid participant or cohort overlap between development and validation. Preprocessing and feature selection should occur within the development process rather than using information from the final test set.

For treatment-resistant depression, direct comparisons with treatment-responsive illness are especially useful, provided that chronicity and current symptoms are also addressed. For schizophrenia, symptom, cognitive, and functioning targets should be separated. These designs would make it possible to determine whether a measure indexes state, accumulated burden, treatment susceptibility, or another clinically relevant dimension.

4.10 Limitations of this review

Retrieval was structured and targeted rather than exhaustive and independently screened. The broad query yield cannot be represented as a PRISMA flow, and publication selection may omit relevant studies. The reference set emphasizes influential syntheses and clinically informative examples rather than every available MRI report.

Extraction relied on verified PubMed records and published summaries. Complete raw imaging data, participant-level clinical information, and all methodological supplements were not available. This

limits detailed appraisal of preprocessing, confounder control, coordinate selection, and outcome multiplicity. Formal risk-of-bias and certainty ratings were not assigned.

The review includes general major-depression evidence to contextualize direct resistance-defined findings. These contextual results cannot be transferred numerically to treatment-resistant depression. Similarly, collaborative schizophrenia and depression cohorts may include mixed ages and diverse clinical states, limiting interpretation as a single uniform adult population.

Many selected publications share cohorts or primary reports. Their sample sizes are not independent totals, and their results were not pooled again. Structural differences, functional convergence, symptom correlations, and treatment predictions are also statistically incompatible without a much narrower original research question.

The figures are descriptive summaries of published effects and a conceptual validation framework. The structural plot omits uncertainty intervals where they were not available in the verified summary. It should not be used to compare diagnostic precision or to infer that one disorder is more severe than another because its group effect is larger.

These limitations mean that the review identifies research candidates and evaluates biomarker claims rather than validating a clinical test. Its main contribution is to preserve the boundaries between group neurobiology, severity association, prognosis, and treatment response, which are essential for responsible translation of MRI findings.

5 CONCLUSION

Hippocampal and prefrontal MRI alterations are credible components of the neurobiology of schizophrenia and depressive illness. Large collaborative analyses support reproducible structural group differences, while functional studies implicate distributed cognitive and affective circuits. Direct treatment-resistant-depression syntheses identify relevant abnormalities but remain constrained by heterogeneous definitions, comparisons, and methods.

These findings do not establish a universally valid individual severity marker. A case-control difference is distinct from a relationship with current symptoms, and treatment-response prediction is distinct from general prognosis. Contextual major-depression effects should not be presented as resistance-specific estimates. Medication, illness course, motion, and cohort overlap further limit straightforward clinical translation.

The most promising path is target-specific validation: define the clinical outcome, establish reliable measurement, test independent cohorts, quantify added value beyond clinical assessment, and evaluate whether marker-informed care improves decisions or outcomes. Hippocampal and prefrontal measures should presently be interpreted as research candidates whose utility depends on the particular clinical task being tested.

Data and methodological transparency

All references correspond to verified PubMed-indexed publications. Numerical figures display reported source estimates; schematic figures summarize concepts and are labeled accordingly. No patient-level dataset, original coordinate-based analysis, new pooled severity correlation, or validated imaging threshold is claimed. The manuscript does not assert a research registration, ethics approval, author funding declaration, or conflict-of-interest statement.

REFERENCES

- Carter LE et al.. Structural and functional brain correlates of treatment-resistant depression: A systematic review of magnetic resonance imaging studies. *Journal of affective disorders*, 409, 121916, 2026. DOI: 10.1016/j.jad.2026.121916. PubMed PMID 42097177
- Kovel CGF et al.. No Alterations of Brain Structural Asymmetry in Major Depressive Disorder: Na ENIGMA Consortium Analysis. *The American journal of psychiatry*, 176(12), 1039-1049, 2019. DOI: 10.1176/appi.ajp.2019.18101144. PubMed PMID 31352813
- Fox MD et al.. Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. *Biological psychiatry*, 72(7), 595-603, 2012. DOI: 10.1016/j.biopsych.2012.04.028. PubMed PMID 22658708
- Frodl T et al.. Childhood adversity impacts on brain subcortical structures relevant to depression. *Journal of psychiatric research*, 86, 58-65, 2017. DOI: 10.1016/j.jpsychires.2016.11.010. PubMed PMID 27918926
- Han LKM et al.. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Molecular psychiatry*, 26(9), 5124-5139, 2021. DOI: 10.1038/s41380-020-0754-0. PubMed PMID 32424236
- Ho TC et al.. Subcortical shape alterations in major depressive disorder: Findings from the ENIGMA major depressive disorder working group. *Human brain mapping*, 43(1), 341-351, 2022. DOI: 10.1002/hbm.24988. PubMed PMID 32198905
- Kaiser RH et al.. Large-Scale Network Dysfunction in Major Depressive Disorder: A Meta-analysis of Resting-State Functional Connectivity. *JAMA psychiatry*, 72(6), 603-11, 2015. DOI: 10.1001/jamapsychiatry.2015.0071. PubMed PMID 25785575

Klok MPC et al.. Structural brain characteristics in treatment-resistant depression: review of magnetic resonance imaging studies. *BJPsych open*, 5(5), e76, 2019. DOI: 10.1192/bjo.2019.58. PubMed PMID 31474243

Leerssen J et al.. Brain structural correlates of insomnia severity in 1053 individuals with major depressive disorder: results from the ENIGMA MDD Working Group. *Translational psychiatry*, 10(1), 425, 2020. DOI: 10.1038/s41398-020-01109-5. PubMed PMID 33293520

Minzenberg MJ et al.. Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Archives of general psychiatry*, 66(8), 811-22, 2009. DOI: 10.1001/archgenpsychiatry.2009.91. PubMed PMID 19652121

Miola A et al.. Structural and functional features of treatment-resistant depression: A systematic review and exploratory coordinate-based meta-analysis of neuroimaging studies. *Psychiatry and clinical neurosciences*, 77(5), 252-263, 2023. DOI: 10.1111/pcn.13530. PubMed PMID 36641802

Schmaal L et al.. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Molecular psychiatry*, 21(6), 806-12, 2016. DOI: 10.1038/mp.2015.69. PubMed PMID 26122586

Schmaal L et al.. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Molecular psychiatry*, 22(6), 900-909, 2017. DOI: 10.1038/mp.2016.60. PubMed PMID 27137745

Shah PJ et al.. Cortical grey matter reductions associated with treatment-resistant chronic unipolar depression. Controlled magnetic resonance imaging study. *The British journal of psychiatry : the journal of mental science*, 172, 527-32, 1998. DOI: 10.1192/bjp.172.6.527. PubMed PMID 9828995

Sun Y et al.. Hippocampal subfield alterations in schizophrenia and major depressive disorder: a systematic review and network meta-analysis of anatomic MRI studies. *Journal of psychiatry & neuroscience : JPN*, 48(1), E34-E49, 2023. DOI: 10.1503/jpn.220086. PubMed PMID 36750240

van Erp TG et al.. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Molecular psychiatry*, 21(4), 547-53, 2016. DOI: 10.1038/mp.2015.63. PubMed PMID 26033243

van Erp TGM et al.. Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biological psychiatry*, 84(9), 644-654, 2018. DOI: 10.1016/j.biopsych.2018.04.023. PubMed PMID 29960671

Weigand A et al.. Prospective Validation That Subgenual Connectivity Predicts Antidepressant Efficacy of Transcranial Magnetic Stimulation Sites. *Biological psychiatry*, 84(1), 28-37, 2018. DOI: 10.1016/j.biopsych.2017.10.028. PubMed PMID 29274805

Wojtalik JA et al.. A Systematic and Meta-analytic Review of Neural Correlates of Functional Outcome in Schizophrenia. *Schizophrenia bulletin*, 43(6), 1329-1347, 2017. DOI: 10.1093/schbul/sbx008. PubMed PMID 28204755

Yun JY et al.. Neural correlates of treatment response to ketamine for treatment-resistant depression: A systematic review of MRI-based studies. *Psychiatry research*, 340, 116092, 2024. DOI: 10.1016/j.psychres.2024.116092. PubMed PMID 39116687