Review Article

Promiscuity of peripheral molecular biomarkers in major psychiatric disorders: a transdiagnostic systematic review

Jairo V. Pinto[1], Thiago C. Moulin[2], Marcia Kauer-Sant'Anna[1], Flávio Kapczinski[1], Olavo B. Amaral[2]

[1]*Department of Psychiatry, Federal University of Rio Grande do Sul. Porto Alegre, RS, Brazil.*

[2]*Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro. Rio de Janeiro, RJ, Brazil.*

Corresponding author:

Olavo B. Amaral
Instituto de Bioquímica Médica
Universidade Federal do Rio de Janeiro
Cidade Universitária, 21941-590
Rio de Janeiro, RJ, Brazil
Phone: (21) 2562-6789
Fax: (21) 2270-8647
E-mail: olavo@bioqmed.ufrj.br

ABSTRACT

The search for biomarkers has been one of the leading endeavours in biological psychiatry; nevertheless, in spite of hundreds of publications, hardly any marker has proved useful in clinical practice. To study how biomarker research has progressed over the years, we performed a systematic review of the literature to evaluate (a) the most studied peripheral molecular markers in major psychiatric disorders, (b) the main experimental design features of studies in which they are proposed as biomarkers and (c) whether their patterns of variation are similar across disorders. An automated search revealed that, out of the six molecules most commonly present as keywords in articles studying plasmatic markers of schizophrenia, major depressive disorder or bipolar disorder, five (BDNF, TNF-alpha, IL-6, C-reactive protein and cortisol) were the same across the three diagnoses. An analysis of the literature on these molecules showed that, whilst 66% of original articles compared their levels between patients and controls, only 35% were longitudinal studies, and only 10% presented an evaluation of diagnostic efficacy, a pattern that has not changed significantly over two decades. Interestingly, these molecules varied similarly across the three disorders, suggesting them to be nonspecific systemic consequences of psychiatric illness rather than diagnostic markers. On the basis of this, we discuss how research fragmentation between diagnoses and publication practices rewarding positive findings may be directing the biomarker literature to nonspecific targets, and what steps could be taken to increase clinical translation in the field.

INTRODUCTION

With the growth of biological psychiatry and the widespread adoption of diagnostic classifications, the concept of diagnostic biomarkers has loomed promisingly in the horizon as one of the major goals in biological psychiatry.[1–3] Although the idea of objective diagnostic tests in psychiatry is not new, and goes back to early promises such as the dexamethasone suppression test,[4] interest in biomarkers has grown exponentially over the last two decades, as shown by the steep rise in the number of articles including the words "biomarker" and "psychiatry" (Figure 1A). Not only the scientific literature, but also news pieces in scientific journals[5] and in the lay media[6] have spread the promise that objective tests might soon trump the centuries-old method of using clusters of symptoms to diagnose mental illness.

Nevertheless, despite years of intensive research, the application of biomarkers in the psychiatric clinic is still very limited, casting scepticism on the idea of replacing symptoms with tests for diagnosis any time soon. As recently discussed by various authors [3,7], problems for translating research findings into clinical practice include the biological heterogeneity of diagnostic constructs,[8–10] the emphasis on statistical rather than clinical significance of findings,[11,12] the emphasis on extreme comparisons between prototypical patients and healthy controls[3,7] and issues such as low statistical power, publication bias and lack of replication among studies.[13] Moreover, the controversy over the soundness of the DSM as a framework for biological psychiatry,[14] as well as the evidence showing that major psychiatric disorders are promiscuous in terms of genetic loci,[15] risk factors,[16] anatomical substrates[17] and treatment,[18] have led some to argue that the greatest promise in

biomarker development might lie not in diagnosis, but in the prediction of prognosis and/or treatment response.[19]

Although such advice seems generally sound, it is unclear whether such concerns have had a significant impact on biomarker research over the years. Moreover, due to the high degree of separation of psychiatric research into different disorders, the question of whether the promiscuity among diagnoses observed for genetics, anatomy and risk factors also applies to peripheral biomarkers has only been studied for isolated cases.[20,21] With this in mind, we decided to (a) systematically investigate what are the most studied plasmatic markers for major psychiatric disorders, (b) perform a systematic review of the experimental design features of articles evaluating them as biomarkers and (c) on the basis of published meta-analyses, investigate whether the variation in their levels is similar across different diagnoses. Our results shed light on current limitations of biomarker research, suggesting that these limitations might have contributed to select nonspecific systemic markers of psychiatric distress as preferential targets for research.

MATERIALS AND METHODS

Search strategy and biomarker selection

We initially searched the PubMed and Scopus databases on June 5th, 2015 for articles containing the terms “biomarker” OR “biomarkers” AND (“serum” OR “blood” OR “plasma” OR “plasmatic”) AND one of six psychiatric disorders: (a) “bipolar disorder”, (b) “major depression” OR “major depressive disorder”, (c) “schizophrenia”, (d) “post-traumatic stress disorder” OR “PTSD”, (e) “attention-deficit-hyperactivity disorder” OR “ADHD”, (f) “autism” OR “ASD”. Although this

search was non-exhaustive, as articles might study peripheral markers without using the term "biomarker", our objective was to specifically select papers on this topic for an automated keyword search – thus, the main goal was to minimize the presence of literature not relating to the theme.

Based on the search results, we used a MATLAB script (available upon request) to count the numbers of articles containing each individual term in the author and index keywords of the Scopus database (the current PubMed XML field for keywords, "Other Terms", was not present for articles published before 2013, and thus could not be used). We then ranked words in order of frequency and manually scanned the tables in order to build a list of the molecules most frequently included as keywords for each individual disorder (Table 1). For this purpose, we took care to aggregate counts of terms relating to the same molecule (e.g. "BDNF" and "brain-derived neutrophic factor").

Selection criteria and analysis of article features

Using the three disorders with the largest number of articles in our search (major depressive disorder, schizophrenia and bipolar disorder), we chose five molecules appearing among the top six keywords for these disorders – brain-derived neurotrophic factor (BDNF), tumor necrosis factor alpha (TNF-α), interleukin-6 (IL-6), C-reactive protein (CRP) and hydrocortisone (cortisol) for further study. We then performed searches in PubMed and Scopus on August 31st, 2015 using the following terms as keywords (as well as MeSH terms for all of them in PubMed):

(a) one of the three disorders ("schizophrenia" / "bipolar disorder" / "major depression" OR "major depressive disorder")

(b) one of the biomarkers above, with inflammatory markers being pooled into a single search due to the high crossover of articles among them ("BDNF" / "interleukin-6" OR "c-reactive protein" OR "tumor necrosis factor alpha" / "hydrocortisone").

(c) "biomarker" OR "biomarkers"

For each biomarker/disorder combination, we screened abstracts (or full text when necessary) for the following inclusion criteria: (a) original articles, (b) in English, (c) including human patients (d) with the disorder of interest, according to DSM or ICD criteria and (e) performing peripheral measurements of the molecule in question (for a summary of the search strategy, see Figure 2). We focused on peripheral levels of the molecules in question, and thus did not include articles evaluating genetic polymorphisms and/or epigenetic changes in their corresponding genes. We then obtained the full text of every article (except for five articles for which it could not be obtained) and extracted the following information on experimental design features (presented in Table 2, Figure 3, Supplementary Tables 1-3 and Supplementary Figure 1):

(a) the specific molecule measured (e.g. BDNF protein, proBDNF protein, BDNF mRNA), site of measurement (e.g. blood, saliva, sweat, urine, hair) and any specific conditions of measurement (e.g. after dexamethasone challenge).

(b) the group comparisons performed in the article (e.g. patients vs. controls; different states of the disorder; patients with the disorder vs. those with another disorder; before vs. after treatment).

(c) the correlations performed within the group of patients (e.g. with symptoms, with illness progression, with treatment response or prognosis, with other markers).

(d) whether the study was cross-sectional or longitudinal.

(e) for longitudinal studies, whether it belonged to a predictive study in a population without the disorder.

(f) whether it offered an objective measure of diagnostic/prognostic efficacy (e.g. receiver operating characteristic (ROC) curves, odds ratio, classifier accuracy) or only presented a statistical comparison between groups.

(g) the total number of markers studied in the article.

(h) the total sample size of the article.

(i) the journal in which it was published.

Information was extracted by two of the authors (O.B.A. and J.V.P.) after extensive discussion of criteria, and 20% of the articles were cross-checked by both investigators, yielding a kappa coefficient of 94.8% for categorical variables. Controversies were solved with the participation of a third investigator (T.C.M.).

Percentages of articles with or without each experimental design feature were initially calculated for each disorder/marker combination. We then calculated aggregate percentages for each disorder or marker and for the whole sample of articles; in these cases, articles appearing in more than one search were counted only once. Articles in which a particular feature was present for one marker/disorder but not for other(s) were considered to include that feature when calculating aggregate percentages. The complete database of articles retrieved by the search, as well as their categorization for each experimental feature, is available as Supplementary Data.

Meta-analysis search and biomarker evaluation across disorders

In order to study whether variation in biomarker levels is similar or distinct among the three analysed disorders, we built descriptive tables summarizing the main

features of each diagnostic marker in each disorder on the basis of existing meta-analyses (Table 3 and Supplementary tables 4-8). For this purpose, we performed additional searches using combinations of markers and disorders in PubMed to locate meta-analyses not included in our initial search, with the last search performed on August 25th, 2016. For supplementary table fields in which no meta-analyses were located, we included summarized results of all articles studying that particular feature in our systematic review sample, but did not search for additional original articles.

Statistical analysis

For comparisons of categorical experimental design features between articles on different markers or different disorders, we used chi-square tests for omnibus comparisons followed by Fisher's exact test between specific pairs of markers/disorders. For comparisons of quantitative features (i.e. sample size and number of markers) between articles on different markers or disorders, we used Kruskal-Wallis tests with Dunn's test as a post-hoc. For correlations of experimental design features of articles with the year of publication, we used rank-biserial correlations for categorical variables and Spearman's nonparametric correlations for quantitative variables (as both year and sample size/number of markers presented heavily skewed distributions).

RESULTS

Our automated keyword search (see flowchart in Figure 2) revealed that, out of the six molecules most commonly present as keywords in articles retrieved using "biomarker" and schizophrenia, major depressive disorder or bipolar disorder, five

(BDNF, TNF-alpha, IL-6, C-reactive protein and cortisol) were the same across these disorders (Table 1). The temporal distribution of these articles (as well as of those fulfilling criteria for subsequent analysis) is displayed on Figure 1 (B and C), revealing that articles on BDNF and inflammatory markers increased sharply after around 2005, whereas the number of articles on cortisol has remained relatively stable over the last two decades. Most articles on BDNF and inflammatory markers studied serum levels of these proteins, while cortisol articles mostly studied serum and salivary cortisol levels, with frequent use of pharmacological challenges such as the dexamethasone suppression test as well (Supplementary Table 1).

An analysis of experimental design features of the literature on these molecules (Table 2 and Supplementary Table 2) showed that, whilst 66% of articles performed comparisons between patients and healthy controls, only 35% were longitudinal studies, and only 10% presented an objective measure of diagnostic or prognostic efficacy (types of measures are detailed in Supplementary Table 3). Most of these numbers did not vary significantly across markers or disorders, but some differences were observed: (a) state comparisons and comparisons with other disorders were more frequent for bipolar disorder, (b) pre vs. post-treatment comparisons and correlations of levels with symptoms were more frequent for BDNF and less frequent for cortisol and (c) cortisol studies presented patient vs. control comparisons less frequently, but included measures of diagnostic efficacy much more frequently than those on other markers (especially inflammation). Median sample size was 71.5 (interquartile range, 45.2-140.8) and larger for inflammatory markers than for BDNF or cortisol ($p=0.004$, Kruskal-Wallis test). The median number of markers addressed in each study was 3 (interquartile range, 1-5), and also larger for studies on inflammation ($p<10^{-4}$, Kruskal-Wallis test).

Temporal trends for the frequency of various types of comparison, as well as for median sample size and number of markers, are shown on Figure 3 and on Supplementary Figure 1. No major differences were observed for the frequency of various types of comparisons or correlations over time, and rank-biserial correlations testing for a time-related trend did not yield p values under 0.05 except for a slight decrease in the frequency of between-disorder comparisons over time (ρ=-0.13, p=0.025). Nevertheless, increases in the total number of articles meant that absolute numbers of articles with any given type of comparison or correlation tended to increase for most categories. Sample size also tended to increase over time (ρ=0.21, $p=4x10^{-4}$), although its distribution varied widely in all periods.

We then moved on to analyse the variation of the five markers in each of the three disorders. Table 3 shows the results of retrieved meta-analyses of patient vs. control comparisons (in various states of the disorder when available) for each marker/disorder combination. Interestingly, one can see that the direction of variation is the same across disorders for all markers (i.e. reductions in BDNF, increases in inflammatory markers and cortisol). In mood disorders, differences between patients and controls tend to be more marked for acutely ill patients (especially for mania in the case of bipolar disorder), while in schizophrenia alterations are generally observed both for acute and chronic illness. This, coupled with the fact that in many cases markers seemed to respond to treatment, especially in the case of BDNF and IL-6 (Supplementary Tables 4-8), suggests these molecules to be nonspecific state markers of psychiatric illness rather than diagnostic markers.

More information on other kinds of comparisons (e.g. correlations with symptoms, illness progression, prediction of treatment response, etc.) can be observed for each marker in Supplementary Tables 4-8. It should be noted, however,

that meta-analyses for such comparisons/correlations were not always found, and that results of individual studies were frequently contradictory. This might be related to differences in the clinical samples studied, but also to the low reliability of these secondary analyses in most articles, due to outcome reporting bias, low statistical power and other features (see Discussion). One should also note that our search strategy was not planned to be exhaustive (except in the case of meta-analyses), and that these tables should not be taken to include the full range of available literature on the subject.

DISCUSSION

Although the topic of biomarker promiscuity has received relatively little attention in the literature, an automated keyword search revealed that the most frequently studied peripheral biomarkers are generally the same across major psychiatric disorders. Besides the biomarkers we chose to focus on (BDNF, IL-6, TNF-α, CRP and cortisol), other molecules were also frequently studied in many disorders, such as oxidative stress markers (glutathione), other cytokines (IL-10, IL-1β) and the astrocytic protein S100B (Table 1). This is not in itself surprising, as one might expect that literature trends will promote interest in similar molecules across different diagnoses. However, the fact that variation patterns in the levels of these markers were also similar across disorders suggests that there are real biological commonalities among them.

The reason for this similarity is worthy of discussion. On one hand, it can be thought of as a sign of overlap between the pathophysiology of major psychiatric disorders, a fact that is also suggested by genetic,[15] risk factor[16] and

neuroanatomical[17] promiscuity among diagnoses. After all, symptom-defined diagnostic boundaries in psychiatry should not be expected to "carve nature at its joints" all the way to the molecular level.[8,44] An additional interesting finding, though, is that many of these markers, such as cytokines and cortisol, are well known to be increased in various forms of chronic stress in both animals and humans.[45] This, along with the fact that these molecules seem to behave as state rather than trait markers in psychiatric disorders, suggests that they might be related to the general "wear and tear" or allostatic load associated with mental or clinical illness,[46] rather than to the specific pathophysiology of any given disorder.

Also in favour of this view is the fact that alterations in cytokines and cortisol can be observed in acute stress models even in normal volunteers.[47–49] And although a BDNF response to acute stress in humans has not been shown, there is evidence that its levels may be altered by traumatic life experiences in psychiatric patients,[50] as well as by acute and chronic stress in animals.[51] The fact that these molecules are altered as a consequence of stress does not preclude, of course, that BDNF decreases, inflammation or hypercortisolemia may also play a causal role in the development and/or progression of mental disorders;[46] however, it does suggest that they are likely to be nonspecific as diagnostic biomarkers.

Interestingly, alterations in both cytokines and cortisol levels have also been found in non-psychiatric medical conditions, such as coronary heart disease[52–54] and cancer.[55,56]. This means that low-grade inflammation related to the consequences of psychiatric illness could be a link between general trends for increased clinical morbidity and mortality in patients with mood disorders and schizophrenia, particularly due to cardiovascular causes.[57,58] Importantly, systemic inflammation can

also have repercussions in the brain, and could be part of a feedback loop that leads to persistence of symptoms in disorders such as major depression.[59,60]

Thus, although the use of the term "biomarker" has become prevalent in psychiatry, and is frequently used in a diagnostic sense, most of the literature using the term seems to focus on nonspecific markers of general distress that are unlikely to be useful for this purpose. This is not to say that these markers are devoid of other clinical applications, as there are multiple potential applications for biomarkers besides diagnosis;[19] the case has been made, for example, to use BDNF and cytokine levels as markers of disease state[20] or progression.[61] However, the fact that the literature on biomarkers is still very fragmented across DSM-defined disorders (only 10% of articles in our sample compared markers across more than one disorder, and this percentage has actually decreased over time) probably limits the understanding of these molecules in this sense, and seems to be an argument in favour of transdiagnostic approaches for their study.[3,14]

The fact that the literature has mostly focused on nonspecific biomarkers also leads to the question of why this has happened – as one might expect that researchers would eventually lose interest in markers that have been shown to be promiscuous. However, this has not been the case, as interest in BDNF and inflammatory markers has only grown over the years, in spite of accumulating evidence of their lack of specificity. Nevertheless, such a trend might be a natural consequence of current incentive systems in science, in which the chase for statistically significant differences and "positive" findings will lead researchers to the areas in which these are most frequently found.[62] Thus, focusing on nonspecific, highly labile markers might be the easiest way to find a significant difference between psychiatric patients and controls, independently of the disorder in study.

Importantly, this also seems to have driven the field to types of studies with less clinical application. The vast majority of articles in our sample focused on comparisons between patients and healthy controls, which are probably not representative of the typical situation in which a biomarker would be useful.[7,12] Also of note is the predominance of cross-sectional studies – which, although necessary as starting points for further research, will not by themselves add diagnostically relevant information in a field where the psychiatric interview is already the gold standard for diagnosis.[63] Another remarkable fact is that measures of diagnostic efficacy that quantify how much a biomarker adds to clinical reasoning, such as ROC curves or odds ratios, were very infrequent in our sample – in fact, they have been *less* reported for recent biomarkers than they were for the evaluation of the dexamethasone suppression test three decades ago.[4]

In a more qualitative note, studies varied widely in terms of methodology, sample size and quality of reporting. Particularly notable were frequent discrepancies between the numbers of measured variables and/or proposed analyses in the methods sections and those reported as results. It was not infrequent for articles to examine a large number of biomarkers, correlate them with a large number of clinical variables, and report only significant associations, in a practice best described as selective outcome reporting bias.[64] The combination of a large number of measured outcomes with selective reporting will inevitably increase the possibility that reported findings are false positives,[13] especially in the absence of multiple comparison corrections, which were infrequently performed in our sample.

Meta-analysis of available data can solve some of these issues, and meta-analytic comparisons have shown that variations in the assessed biomarkers seem to be robust in common psychiatric disorders (Table 3). However, meta-analyses are

intrinsically limited by the quality of the data; thus, their results can be influenced by publication and outcome reporting biases in the literature, as recently studied for markers of bipolar disorder.[65] Moreover, they have mostly been performed for comparisons that are prevalent among studies – i.e. those between patients and controls, and eventually between disease states or pre- and post-treatment levels. Attempts to correlate biomarkers with symptom severity, illness progression or response to treatment, on the other hand, have yielded less consistent findings in meta-analyses (with a few exceptions), probably due to the fact that these correlations are not only less frequent among articles, but also usually presented as secondary analyses, and thus more susceptible to bias.

The current picture of the peripheral biomarker literature, in this sense, is reminiscent of the early days of genetic epidemiology studies, in which a large number of underpowered studies yielded a multitude of gene-disease associations with inflated effect sizes and limited reproducibility.[66,67] That field has since moved on to much bigger, adequately powered studies performed by multicentre consortia[68] that have been able to control for the large number of comparisons performed and to yield more reproducible data.[69] However, this does not seem to be occurring in the peripheral biomarker literature yet. On the contrary, aside from relatively weak trends for an increase in sample size and a decrease in the frequency of cross-disorder comparisons, hardly any of the literature patterns observed seems to have changed significantly over a 20-year period.

Such a general overview of the literature as performed in our review naturally has its limitations. First and foremost, the use of the term “biomarker” in our searches was a deliberate choice for specificity over sensitivity, and probably led many studies referring to peripheral markers by other terms to be missed. It might also have biased

our sample to particular types of study, although we cannot say if that was the case. Moreover, in some articles the markers we assessed were only one among many studied, and not necessarily the main focus of the work – thus, the prevalence of some analyses or comparisons might have been larger if we had limited ourselves to studies focused on those specific markers. Nevertheless, we chose not to exclude those studies, as lack of reporting of all the analyses performed in a study is in itself a problem. Finally, our choice to provide a bird's eye view of the literature has led us to focus on the rule rather than on the exceptions – thus, one should keep in mind that there are several articles in our sample that ask (and eventually answer) meaningful clinical questions. Moreover, although in terms of prevalence they constitute a minority, the absolute number of these articles tends to increase as the biomarker literature grows as a whole.

In spite of these limitations, we believe that our general conclusions concerning the psychiatric peripheral biomarker literature – namely, that the most frequently studied markers are nonspecific state markers for multiple disorders, and that the comparisons performed in most articles are not sufficient to generate useful clinical information – are probably correct. Therefore, it is useful to discuss ways in which this picture can be improved. First and foremost, it is important to define what one means by "biomarker", as there are many ways in which a molecule might behave as a clinically useful marker.[19] Moreover, we must move from studies chasing simple differences between groups to those that try to measure how much difference a biomarker can make in a clinical decision process.[3,7,12] Finally, more meaningful studies will likely require more rigid design and larger statistical power – in this sense, development and use of reporting guidelines for biomarker studies,[70] pre-registration of study protocols to avoid data dredging[71] and formation of multicentre

consortia using more standardised research designs[68] are foreseeable ways to improve reproducibility and clinical translation in the field.

Still, it's important to note that, while such suggestions have been in place for years, they have not made a major impact on research in the field up to now. Although one might have hoped that the optimism generated by initial findings of biomarker alterations in psychiatric patients would have been followed by larger studies investigating these in more detail – as occurred with the dexamethasone suppression test, for example,[4] this does not seem to have happened. On the contrary, these initial descriptions seem to have led to ever more numerous articles measuring an increasing number of biomarkers in similar situations (e.g. simple patient-control comparisons). It is likely that these patterns will only change if research incentives shift from rewarding publication to rewarding reproducible and clinically useful findings, a point that probably holds true for other fields of science as well.[72,73]

Finally, one has to consider that, if more rigorous standards are applied to the peripheral biomarker literature, it is possible that the field will reach the conclusion that many of the currently studied molecules might not be particularly useful in clinical settings, especially for diagnostic purposes, as they seem to be markers of general distress rather than of specific symptoms or syndromes. It might be expected, after all, that peripherally measured molecules will be limited in their correlations with symptoms that are produced by the highly specialised anatomy of brain circuits. Moreover, the fact that most single molecules have very modest impacts in the development of psychiatric illness (as expected for a complex system such as the brain) has been made clear by genomic studies,[74] and is likely to hold true for peripherally measured molecules as well.

Even if this is the case, however, such research will not have been in vain, as it might shed light on the connections between psychiatric and medical illness, yield insights on pathophysiology, and perhaps provide ways to assess risk, disease progression and/or severity, especially if various markers are used in concert.[61,75] Moreover, it can also increase our knowledge of the consequences of chronic stress on the brain and body, and help in the creation of transdiagnostic approaches to bridge the gaps between psychiatric research and neuroscience, or between psychiatry and other fields of medicine.[14] For this to happen, however, the field needs to taper down its initial optimism, acknowledge that most markers will not be diagnostic or specific, increase the rigour of its approaches and focus on meaningful clinical questions that can drive research from statistical to real-world significance.

AUTHOR CONTRIBUTIONS

O.B.A., M.K.S. and J.V.P. designed the study. J.V.P. performed literature searches. T.C.M. programmed the text-mining algorithm. J.V.P., T.C.M. and O.B.A. extracted article data and cross-checked data for consistency. T.C.M., O.B.A. and J.V.P. designed figures and tables. O.B.A., M.K.S., F.K. and J.V.P. interpreted data. O.B.A. wrote the initial version on the manuscript. All authors provided critical input in revisions of the manuscript.

REFERENCES

1 Kupfer D, First MB, Regier DA. A Research Agenda for DSM-V. American Psychiatric Association, Washington, 2002.

2 Insel TR, Quirion R. Psychiatry as a clinical neuroscience discipline. *JAMA* 2005; **294**: 2221–4.

3 The Lancet Psychiatry TL. Blood biomarkers in psychiatry. *Lancet Psychiatry* 2016; **3**: 693.

4 Arana GW. The dexamethasone suppression test for diagnosis and prognosis in psychiatry. *Arch Gen Psychiatry* 1985; **42**: 1193.

5 Haag AL. Biomarkers trump behavior in mental illness diagnosis. *Nat Med* 2007; **13**: 3.

6 Nutt AE. Brain hacking: The mind's biology. The Washington Post, 2016. http://www.washingtonpost.com/sf/national/2016/02/19/brain-hacking-the-minds-biology/ (accessed May 24, 2016).

7 Kapur S, Phillips a G, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* 2012; **17**: 1174–9.

8 Pavão R, Tort ABL, Amaral OB. Multifactoriality in psychiatric disorders: a computational study of schizophrenia. *Schizophr Bull* 2015; **41**: 980–8.

9 Hyman SE. The diagnosis of mental disorders: The problem of reification. *Annu Rev Clin Psychol* 2010; **6**: 155–79.

10 Heinrichs RW. Meta-analysis and the science of schizophrenia: variant evidence or evidence of variants? *Neurosci Biobehav Rev* 2004; **28**: 379–94.

11 Sterne J, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ* 2001; **322**: 226–31.

12 Perlis RH. Translating biomarkers to clinical practice. *Mol Psychiatry* 2011; **16**: 1076–87.

13 Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005; **2**. DOI:10.1371/journal.pmed.0020124.

14 Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013; **11**: 126.

15 Smoller JW. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* 2013; **381**: 1371–9.

16 Caspi A, Houts RM, Belsky DW, *et al.* The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci* 2014; **2**: 119–37.

17 Goodkind M, Eickhoff SB, Oathes DJ, *et al.* Identification of a common

neurobiological substrate for mental illness. *JAMA Psychiatry* 2015; **72**: 305–15.

18 Chouinard G. The search for new off-label indications for antidepressant, antianxiety, antipsychotic and anticonvulsant drugs. *J Psychiatry Neurosci* 2006; **31**: 168–76.

19 Davis J, Maes M, Andreazza A, McGrath JJ, Tye SJ, Berk M. Towards a classification of biomarkers of neuropsychiatric disease: from encompass to compass. *Mol Psychiatry* 2015; **20**: 152–3.

20 Fernandes BS, Berk M, Turck CW, Steiner J, Gonçalves C. Decreased peripheral brain-derived neurotrophic factor levels are a biomarker of disease activity in major psychiatric disorders: a comparative meta-analysis. *Mol Psychiatry* 2014; **19**: 749–51.

21 Goldsmith DR, Rapaport MH, Miller BJ. A meta-analysis of blood cytokine network alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder and depression. *Mol Psychiatry* 2016; doi: 10.1038/mp.2016.3

22 Polyakova M, Stuke K, Schuemberg K, Mueller K, Schoenknecht P, Schroeter ML. BDNF as a biomarker for successful treatment of mood disorders: A systematic & quantitative meta-analysis. *J Affect Disord* 2015; **174**: 432–40.

23 Fernandes BS, Molendijk ML, Köhler CA, *et al.* Peripheral brain-derived neurotrophic factor (BDNF) as a biomarker in bipolar disorder: a meta-analysis of 52 studies. *BMC Med* 2015; **13**: 289.

24 Munkholm K, Vinberg M, Kessing LV. Peripheral blood brain-derived neurotrophic factor in bipolar disorder: a comprehensive systematic review and meta-analysis. *Mol Psychiatry* 2016; **21**: 216-228.

25 Fernandes BS, Steiner J, Berk M, *et al.* Peripheral brain-derived neurotrophic factor in schizophrenia and the role of antipsychotics: meta-analysis and implications. *Mol Psychiatry* 2015; **20**: 1108-1119.

26 Haapakoski R, Mathieu J, Ebmeier KP, Alenius H, Kivimäki M. Cumulative meta-analysis of interleukins 6 and 1β, tumour necrosis factor α and C-reactive protein in patients with major depressive disorder. *Brain Behav Immun* 2015; **49**: 206–15.

27 Hiles SA, Baker AL, de Malmanche T, Attia J. A meta-analysis of differences in IL-6 and IL-10 between people with and without depression: Exploring the

causes of heterogeneity. *Brain Behav Immun* 2012; **26**: 1180–8.

28 Liu Y, Ho RCM, Mak A. Interleukin (IL)-6, tumour necrosis factor alpha (TNF-α) and soluble interleukin-2 receptors (sIL-2R) are elevated in patients with major depressive disorder: A meta-analysis and meta-regression. *J Affect Disord* 2012; **139**: 230–9.

29 Dowlati Y, Herrmann N, Swardfager W, *et al.* A meta-analysis of cytokines in major depression. *Biol Psychiatry* 2010; **67**: 446–57.

30 Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosom Med* 2009; **71**: 171–86.

31 Modabbernia A, Taslimi S, Brietzke E, Ashrafi M. Cytokine alterations in bipolar disorder: A meta-analysis of 30 studies. *Biol Psychiatry* 2013; **74**: 15–25.

32 Munkholm K, Braüner JV, Kessing LV, Vinberg M. Cytokines in bipolar disorder vs. healthy control subjects: A systematic review and meta-analysis. *J Psychiatr Res* 2013; **47**: 1119–33.

33 Munkholm K, Vinberg M, Kessing LV. Cytokines in bipolar disorder: A systematic review and meta-analysis. *J Affect Disord* 2013; **144**: 16–27.

34 Potvin S, Stip E, Sepehry AA, Gendron A, Bah R, Kouassi E. Inflammatory cytokine alterations in schizophrenia: a systematic quantitative review. *Biol Psychiatry* 2008; **63**: 801–8.

35 Miller BJ, Buckley P, Seabolt W, Mellor A, Kirkpatrick B. Meta-analysis of cytokine alterations in schizophrenia: Clinical status and antipsychotic effects. *Biol Psychiatry* 2011; **70**: 663–71.

36 Upthegrove R, Manzanares-Teson N, Barnes NM. Cytokine function in medication-naive first episode psychosis: A systematic review and meta-analysis. *Schizophr Res* 2014; **155**: 101–8.

37 Dargél AA, Godin O, Kapczinski F, Kupfer DJ, Leboyer M. C-reactive protein alterations in bipolar disorder: A meta-analysis. *J Clin Psychiatry* 2015; **76**: 142–50.

38 Fernandes B, Steiner J, Bernstein HG, *et al.* C-reactive protein is increased in schizophrenia but is not altered by antipsychotics: meta-analysis and implications. *Mol Psychiatry* 2016; **21**: 554–64.

39 Stetler C, Miller GE. Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. *Psychosom*

Med 2011; **73**: 114–26.

40 Lopez-Duran NL, Kovacs M, George CJ. Hypothalamic-pituitary-adrenal axis dysregulation in depressed children and adolescents: A meta-analysis. *Psychoneuroendocrinology* 2009; **34**: 1272–83.

41 Murri MB, Pariante C, Mondelli V, *et al.* HPA axis and aging in depression: Systematic review and meta-analysis. *Psychoneuroendocrinology*. 2014; **41**: 46–62.

42 Girshkin L, Matheson SL, Shepherd AM, Green MJ. Morning cortisol levels in schizophrenia and bipolar disorder: A meta-analysis. *Psychoneuroendocrinology* 2014; **49**: 187–206.

43 Murri MB, Prestia D, Mondelli V, *et al*. The HPA axis in bipolar disorder: Systematic review and meta-analysis. *Psychoneuroendocrinology* 2016; **63**: 327–42.

44 Kendler KS. Reflections on the relationship between psychiatric genetics and psychiatric nosology. *Am J Psychiatry* 2006; **163**: 1138–46.

45 McEwen BS. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol Rev* 2007; **87**: 873–904.

46 Kapczinski F, Vieta E, Andreazza AC, *et al*. Allostatic load in bipolar disorder: Implications for pathophysiology and treatment. *Neurosci Biobehav Rev* 2008; **32**: 675–92.

47 Slavich GM, Way BM, Eisenberger NI, Taylor SE. Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proc Natl Acad Sci USA* 2010; **107**: 14817–22.

48 Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychol Bull* 2004; **130**: 355–91.

49 Steptoe A, Hamer M, Chida Y. The effects of acute psychological stress on circulating inflammatory factors in humans: a review and meta-analysis. *Brain Behav Immun* 2007; **21**: 901–12.

50 Kauer-Sant'Anna M, Tramontina J, Andreazza AC, *et al.* Traumatic life events in bipolar disorder: Impact on BDNF levels and psychopathology. *Bipolar Disord* 2007; **9 (Suppl 1)**: 128–35.

51 Duman RS, Monteggia LM. A neurotrophic model for stress-related mood disorders. *Biol Psychiatry* 2006; **59**: 1116–27.

52 Mendall MA, Patel P, Asante M, *et al.* Relation of serum cytokine concentrations to cardiovascular risk factors and coronary heart disease. *Heart* 1997; **78**: 273–7.

53 Danesh J, Wheeler JG, Hirschfield GM, *et al.* C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *N Engl J Med* 2004; **350**: 1387–97.

54 Smith GD, Ben-Shlomo Y, Beswick A, Yarnell J, Lightman S, Elwood P. Cortisol, testosterone, and coronary heart disease: prospective evidence from the Caerphilly study. *Circulation* 2005; **112**: 332–40.

55 Martín F, Santolaria F, Batista N, *et al.* Cytokine levels (IL-6 and IFNγ), acute phase response and nutritional status as prognostic factors in lung cancer. *Cytokine* 1999; **11**: 80–6.

56 Rich T, Innominato PF, Boerner J, *et al.* Elevated serum cytokines correlated with altered behavior, serum cortisol rhythm, and dampened 24-hour rest-activity patterns in patients with metastatic colorectal cancer. *Clin Cancer Res* 2005; **11**: 1757–64.

57 Olfson M, Gerhard T, Huang C, Crystal S, Stroup TS. Premature mortality among adults with schizophrenia in the united states. *JAMA Psychiatry* 2015; **72**: 1172–81.

58 Ösby U, Westman J, Hällgren J, Gissler M. Mortality trends in cardiovascular causes in schizophrenia, bipolar and unipolar mood disorder in Sweden 1987-2010. *Eur J Public Health* 2016; doi: 10.1093/eurpub/ckv245

59 Wohleb ES, Franklin T, Iwata M, Duman RS. Integrating neuroimmune systems in the neurobiology of depression. *Nat Rev Neurosci* 2016; **17**: 497–511.

60 Gold PW. The organization of the stress system and its dysregulation in depressive illness. *Mol Psychiatry* 2015; **20**: 32-47.

61 Kapczinski F, Dias VV, Kauer-Sant'Anna M, *et al.* The potential use of biomarkers as an adjunctive tool for staging bipolar disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 2009; **33**: 1366–71.

62 Smaldino PE, McElreath R. The natural selection of bad science. 2016. http://arxiv.org/abs/1605.09511 (accessed Aug 21, 2016).

63 Amaral OB. Do biomarkers trump behavior? *Nat Med* 2007; **13**: 237.

64 Dwan K, Altman DG, Arnaiz JA, *et al.* Systematic review of the empirical

evidence of study publication bias and outcome reporting bias. *PLoS One* 2008; **3**: e3081.

65 Carvalho AF, Köhler CA, Fernandes BS, *et al.* Bias in emerging biomarkers for bipolar disorder. *Psychol Med* 2016; **46**: 2287-2297.

66 Bosker FJ, Hartman CA, Nolte IM, *et al.* Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Mol Psychiatry* 2011; **16**: 516–32.

67 Siontis KCM, Patsopoulos NA, Ioannidis JPA. Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *Eur J Hum Genet* 2010; **18**: 832–7.

68 Sullivan PF. The psychiatric GWAS consortium: Big science comes to psychiatry. *Neuron* 2010; **68**: 182–6.

69 Collins AL, Sullivan PF, Corvin A, *et al.* Genome-wide association studies in psychiatry: what have we learned? *Br J Psychiatry* 2013; **202**: 1–4.

70 Gnanapavan S, Hegen H, Khalil M, *et al.* Guidelines for uniform reporting of body fluid biomarker studies in neurologic disorders. *Neurology* 2014; **83**: 1210–6.

71 Kivimaki M, Singh-Manoux A, Ferrie JE, Batty GD. Post hoc decision-making in observational epidemiology--is there need for better research standards? *Int J Epidemiol* 2013; **42**: 367–70.

72 Ioannidis JPA. How to make more published research true. *PLoS Med* 2014; **11:** e1001747.

73 Ioannidis JPA, Greenland S, Hlatky MA, *et al.* Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014; **383**: 166–75.

74 Kendler KS. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Mol Psychiatry* 2013; **18**: 1058–66.

75 Kapczinski F, Dal-Pizzol F, Teixeira a L, *et al.* A systemic toxicity index developed to assess peripheral changes in mood episodes. *Mol Psychiatry* 2010; **15**: 784–6.

Table 1.

MDD		BD		Schizophrenia		PTSD		Autism		ADHD	
Molecule	#	Molecule	#	Molecule	#	Molecule	#	Molecule	#	Molecule	#
BDNF	**86**	**BDNF**	**55**	**BDNF**	**48**	**Cortisol**	**19**	Glutathione	20	*Dopamine*	*7*
Cortisol	**61**	**TNF-α**	**21**	*Prolactin*	*31*	**CRP**	**12**	Serotonin	20	Norepinephrine	5
IL-6	**60**	**IL-6**	**18**	**TNF-α**	**22**	**IL-6**	**11**	**BDNF**	**12**	**BDNF**	**4**
CRP	**53**	**Cortisol**	**13**	**IL-6**	**19**	*CRH*	*7*	Homocysteine	12	Adiponectin	4
TNF-α	**46**	**CRP**	**12**	Glutathione	17	**BDNF**	**6**	**TNF-α**	**10**	**Cortisol**	**3**
Fibrinogen	23	S100B	11	*Insulin*	*17*	ACTH	6	**IL-6**	**9**	*Oxyhemoglobin*	*3*
IL-1 beta	21	SOD	10	*Leptin*	*15*	IL-8	5	Cysteine	8	*Serotonin*	*3*
Serotonin	*18*	*Glutamic acid*	*10*	*Glucose*	*14*	Neuropeptide Y	5	Interferon-γ	8	*Cholesterol*	*2*
IL-10	16	IL-10	8	**Cortisol**	**13**	IL-1 beta	4	Oxytocin	8	**IL-6**	**2**
S100B	16	Neurotrophin-3	8	*Dopamine*	13	IL-4	4	Lactic acid	7	**CRP**	**2**
Glutamic acid	*13*	NGF	7	**CRP**	**12**	Interferon-γ	4	Methionine	7	MHPG	2
Prolactin	*12*	*GABA*	*7*	*GABA*	*12*	**TNF-α**	**4**	Melatonin	7	Ferritin	2
Insulin	*11*	IL-1 beta	6	ACTH	12	Annexin a2	4	*Glutamic acid*	7	Homovanillic acid	2
GABA	*11*	IL-2	6	Interferon-γ	11	Serotonin	4	*GABA*	6	Neuropeptide Y	2
Interferon-γ	10	Nitric oxide	6	Homocysteine	11	S100B	4	IL-1 beta	5	DOPAC	2
Kynurenine	10	*Dopamine*	*6*	IL-10	11	IL-2	3	EGF	5	*Iron blood level*	2
Leptin	*10*	*Insulin*	*6*	*Glutamic acid*	*11*	*Leptin*	*3*	*Various*	-	*Various*	-

Table 1. Most frequent molecules among Scopus keywords in peripheral biomarker studies of different psychiatric disorders. Table shows the top-ranked endogenous molecules among Scopus keywords in a database search for "biomarker" AND "serum OR blood OR plasma OR plasmatic" AND individual disorder (see methods). Note that not all molecules included as keywords necessarily represent candidate peripheral biomarkers. The 5 top hits for mood disorders are in bold, and at least 3 of them appear in every disorder. Italic indicates molecules used as markers of treatment effects (e.g. prolactin, insulin) and molecules not used as peripheral markers, which appeared in the text for other reasons (e.g. glutamic acid, serotonin). ACTH: adenocorticotropic hormone; BDNF: brain-derived neurotrophic factor; CRH: corticotrophin releasing hormone; CRP: C-reactive protein; DOPAC: 3, 4 dihydroxyphenylacetic acid; EGF: epidermal growth factor; GABA: gamma-aminobutyric acid: IL-1 beta: interleukin 1 beta; IL-2: interleukin 2; IL-4: interleukin 4; IL-6: interleukin 6; IL-8: interleukin 8; IL-10: interleukin 10; MHPG: 4-hydroxy-3 methoxyphenylethylene glycol; NGF: nerve growth factor; S100B: S100 calcium-binding protein beta; SOD: superoxide dismutase; TNF-α: tumor necrosis factor alpha.

Table 2.

	MDD				BD				Schizophrenia				p	All
	BDNF	**Inflam**	**Cort**	**Total**	**BDNF**	**Inflam**	**Cort**	**Total**	**BDNF**	**Inflam**	**Cort**	**Total**		
Included/total articles	52/149	82/217	56/152	**182/518**	25/65	27/63	10/34	**56/162**	19/68	32/84	14/34	**60/186**	-	**280/866**
Group comparison														
vs. control group	31 (60%)	60 (73%)	31 (55%)	**117 (64%)**	19 (76%)	19 (70%)	5 (50%)	**38 (68%)**	15 (79%)	23 (72%)	6 (43%)	**42 (70%)**	.687	**184 (66%)**
between disease states	2 (4%)	4 (5%)	0 (0%)	**6 (3%)[2]**	8 (32%)	7 (26%)	3 (30%)	**17 (30%)[1,3]**	2 (11%)	1 (3%)	0 (0%)	**3 (5%)[2]**	$<10^{-4}$	**26 (9%)**
vs. other disorder	9 (17%)	7 (9%)	7 (13%)	**21 (12%)[2]**	7 (28%)	6 (22%)	4 (40%)	**15 (27%)[1]**	3 (16%)	5 (16%)	2 (14%)	**8 (13%)**	.018	**28 (10%)**
before vs. after treatment	24 (46%)	28 (34%)	7 (13%)	**56 (31%)**	7 (28%)	2 (7%)	1 (10%)	**10 (18%)**	2 (11%)	7 (22%)	4 (29%)	**12 (20%)**	.075	**77 (28%)**
Correlation														
w/ current symptoms	24 (46%)	31 (38%)	15 (27%)	**66 (36%)**	12 (48%)	5 (19%)	1 (20%)	**17 (30%)**	10 (53%)	12 (37%)	5 (36%)	**24 (40%)**	.550	**103 (37%)**
w/ progression	5 (10%)	5 (6%)	2 (4%)	**12 (7%)**	3 (12%)	5 (19%)	2 (20%)	**9 (16%)**	2 (11%)	6 (19%)	0 (0%)	**8 (13%)**	.064	**29 (10%)**
w/ treatment response	14 (27%)	12 (15%)	5 (9%)	**29 (16%)**	3 (12%)	3 (11%)	0 (0%)	**6 (11%)**	1 (5%)	4 (12%)	3 (21%)	**8 (13%)**	.601	**42 (15%)**
Other														
Risk prediction	3 (6%)	2 (2%)	4 (7%)	**9 (5%)**	1 (4%)	1 (4%)	1 (10%)	**3 (5%)**	0 (0%)	1 (3%)	0 (0%)	**1 (2%)**	.515	**12 (4%)**
Study features														
Longitudinal	27 (51%)	29 (35%)	14 (25%)	**67 (37%)**	10 (40%)	6 (21%)	4 (40%)	**20 (6%)**	2 (11%)	8 (25%)	5 (36%)	**14 (23%)**	.152	**98 (35%)**
Diagnostic efficacy	8 (15%)	4 (5%)	15 (27%)	**23 (13%)**	3 (12%)	1 (4%)	2 (20%)	**5 (9%)**	2 (11%)	4 (12%)	1 (7%)	**5 (8%)**	.557	**29 (10%)**

Table 2. Experimental design features of retrieved articles for each biomarker/disorder combination. Articles measuring more than one marker for a given disorder or the same marker for more than one disorder were counted only once to calculate totals. BDNF, brain-derived neurotrophic factor; MDD, major depressive disorder; BD: bipolar disorder; SCZ: schizophrenia; p values refer to a χ^2 test comparing aggregate values for all markers between the 3 disorders [1]$p<0.05$ vs. MDD, [2]$p<0.05$ vs. BD, [3] $p<0.05$ vs. SCZ, Fisher's exact test comparing pairs of disorders. A similar table with aggregated totals for individual markers instead of disorders is presented as Supplementary Table 2.

Table 3.

	BDNF	IL-6	TNF-alpha	C-reactive protein	Cortisol
MDD	Depression: ↓[22] Euthymia: =[22]	Overall: ↑[26-30] Acute: ↑[21] Chronic: ↑[21]	Overall: ↑[26,28,29] Acute: ↑[21] Chronic: =[21]	Overall: ↑[26,30]	Overall: ↑[39-41]
BD	Overall: ↓[22-24] Mania: ↓[22-24] Depression: ↓[22-24] Mixed: =[24] Euthymia: =[22-24]	Overall: n.s.↑[31,32] Mania: ↑[21,31], =[33] Depression: =[21,31,33] Euthymia: =[31,33], ↑[21]	Overall: ↑[31,32] Mania: ↑[21,31,33] Depression: n.s.↑[31,33], =[21] Euthymia: =[21,31,33]	Overall: ↑[37] Mania: ↑[37] Depression: =[37] Euthymia: ↑[37]	Overall: ↑[42,43] Mania: ↑[43], =[42] Depression: =[43] Euthymia: ↑[43]
SCZ	Overall: ↓[25] First-episode: ↓[25] Drug naive: ↓[25] Medicated: ↓[25] Chronic: ↓[25]	Overall: ↑[34] First-episode: ↑[21,35,36] Acute: ↑[21,35] Stable/Chronic: ↑[21], =[35]	Overall: =[34] First-episode: ↑[21,35,36] Acute: ↑[21,35] Chronic: ↑[21]	Overall: ↑[38] First-episode: ↑[38] Chronic: ↑[38]	Overall: ↑[42] First-episode: =[42] Chronic: ↑[42]

Table 3. **Summary of variation in serum levels of different markers across disorders.** Arrows indicate the direction of variation (vs. control levels) for each disorder combination: ↑ and ↓ indicate significant variation with $p < 0.05$ (n.s. indicates non-significant trend with $p = 0.05 – 0.1$); = indicates no significant variation (which does not imply that levels are similar, as lack of statistical significance may also be due to low power). All comparisons are based on the most recent meta-analytic estimate(s) available – if there were multiple recent meta-analyses, or if they presented significant methodological differences (e.g. inclusion criteria) among them, more than one was included. For disorder/marker combinations in which meta-analysis results were available for various states of the disorders, these were included as well. MDD; major depressive disorder; BD, bipolar disorder; SCZ, Schizophrenia; BDNF, brain-derived neurotrophic factor; IL-6, interleukin 6; TNF-alpha, tumor necrosis factor alpha. For [43], only morning cortisol levels were considered.

Figure 1.

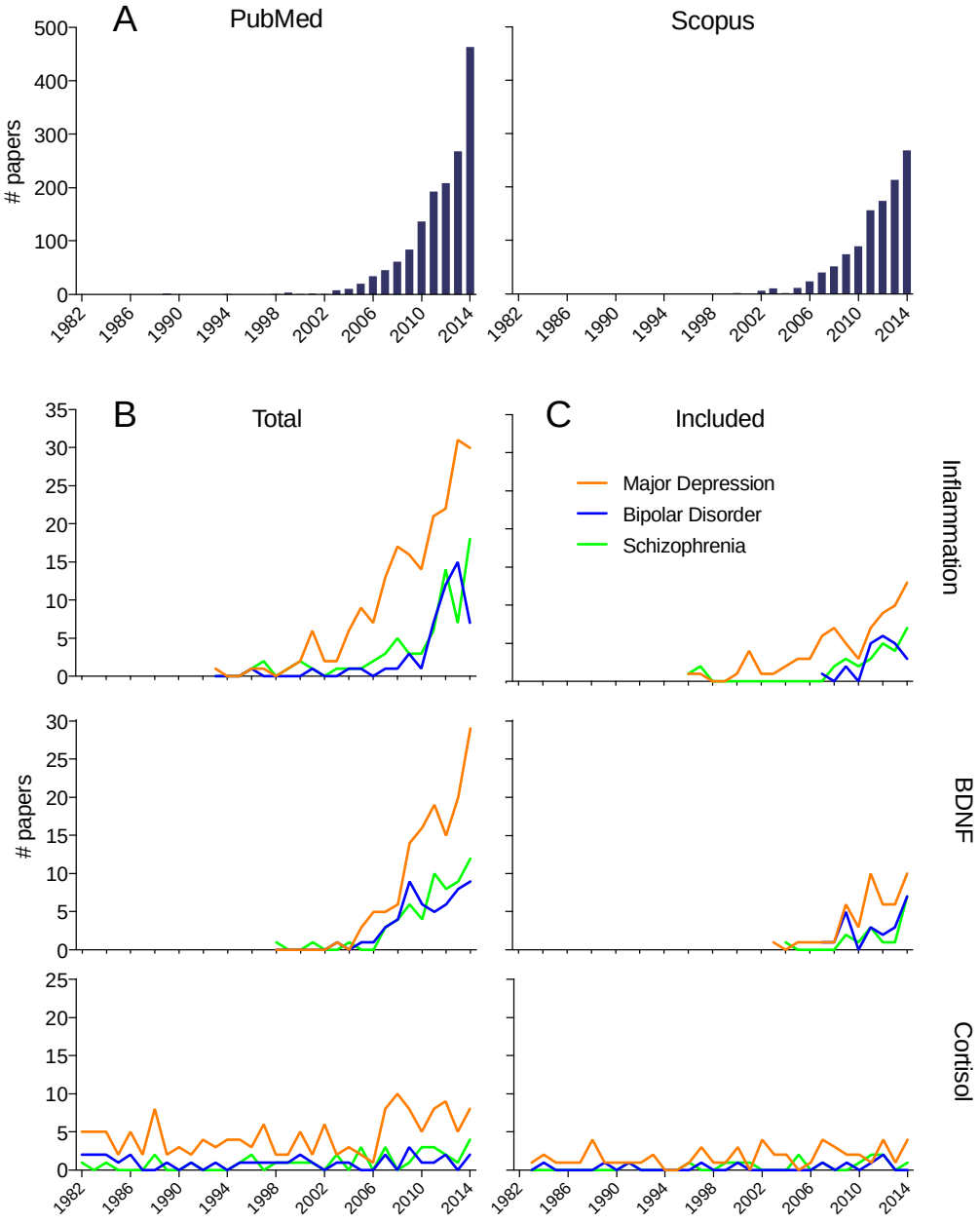

Figure 1. Growth of biomarker research in psychiatry (A) Number of PubMed (left) and Scopus (right) hits in searches for "biomarker" AND "psychiatry" or "psychiatric" for each year between 1982 and 2014 **(B)** Total number of articles over the same period retrieved for searches in PubMed and Scopus for "biomarker" AND either (a) "tumor necrosis factor alpha OR interleukin-6 OR C-reactive protein" (top) or (b) "BDNF" (middle) or (c) "hydrocortisone" (bottom) AND either (a) "major depression OR depressive disorder" (orange) (b) "bipolar disorder" (blue) or (c) "schizophrenia" (using Pubmed MeSH terms); **(C)** original articles fulfilling criteria for inclusion in our systematic review of experimental design features for each year.

Figure 2.

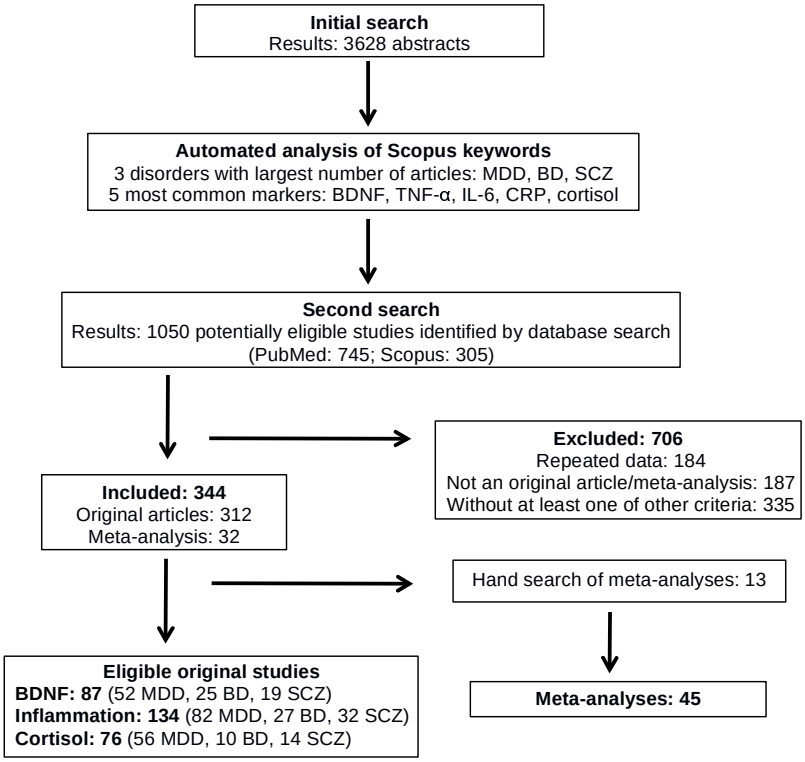

Figure 2. Flowchart depicting the selection of studies. Flowchart detailing the various stages of study selection in the studies. A more detailed account of the procedure can be found in the methods section and in the Supplementary Appendix. Eligible original studies for each biomarker include articles appearing in more than one search – thus, the sum of articles for individual disorders is greater than the total number of articles for each marker. MDD, major depressive disorder; BD, bipolar disorder; SCZ, schizophrenia; BDNF, brain-derived neurotrophic factor; TNF-α, tumor necrosis factor alpha; IL-6, interleukin-6, CRP, C-reactive protein.

Figure 3.

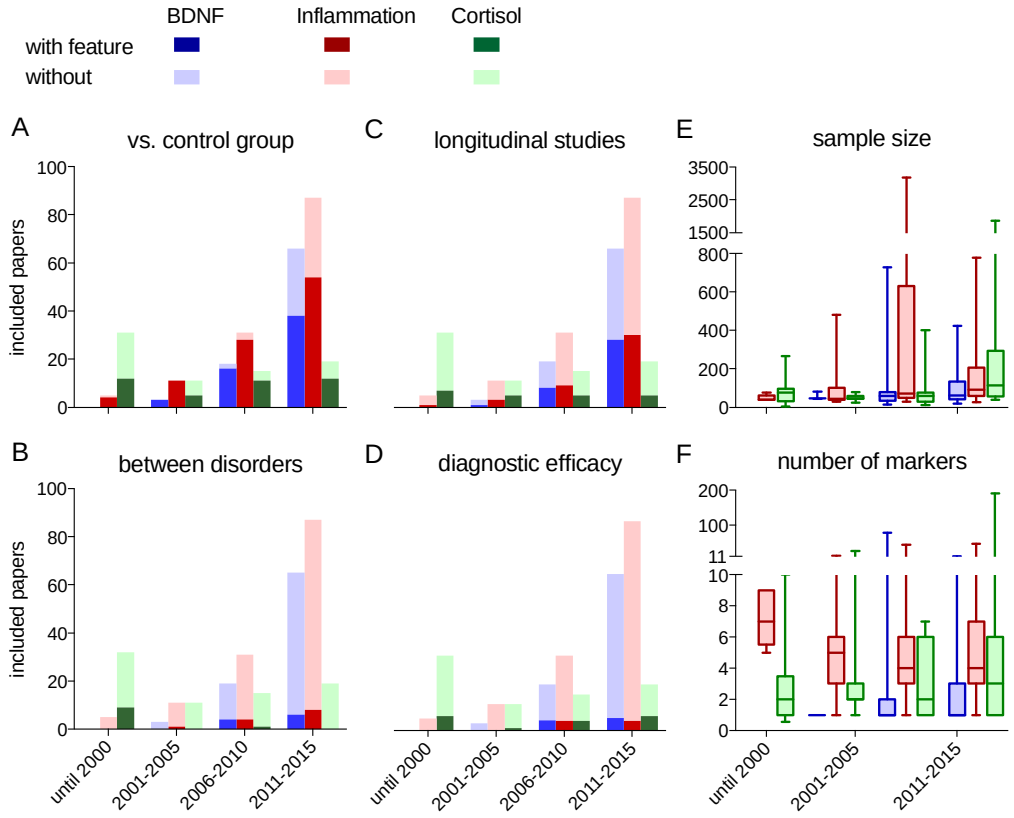

Figure 3. Variations of experimental design features of articles over different periods. Figure shows the frequency of **(A)** patient control-comparisons, **(B)** comparisons between disorders, **(C)** comparisons including a measure of diagnostic efficacy and **(D)** longitudinal studies, as well as the distribution of **(E)** sample size and **(F)** number of peripheral biomarkers studied among analysed articles measuring BDNF (blue), TNF-α/IL-6/CRP (red) and cortisol (green) over 4 epochs (until 2000, 2001-2005, 2006-2010, 2011-2015). For dichotomous variables (A-D), bars represent the total number of articles, with dark shading representing articles with each experimental design feature and light shading indicating those without them. Distributions of quantitative variables (E-F) are expressed as box-whisker plots (center line, median; box, interquartile range, whiskers, $5^{th}/95^{th}$ percentiles). Spearman's ρ coefficients for correlations between year of publication and each feature are (A) ρ=-0.01, p=0.80, (B) ρ=-0.13, p=0.025, (C) ρ=-0.04, p=0.47, (D) ρ=0.08, p=0.20, (E) ρ=0.21, $p=4x10^{-4}$, (F) ρ=-0.003, p=0.95. Variations of additional features can be visualised in Supplementary Figure 1.