

# Measurement Without Understanding: Commercial Incentives, Diagnostic Expansion, and the Over-Diagnosis of Mental Illness

Austin Jones<sup>1</sup>

*Received July 5, 2026*

*Accepted October 9, 2026*

*Electronic access November 15, 2026*

Psychiatric diagnosis has expanded faster than the science beneath it. The Diagnostic and Statistical Manual of Mental Disorders (DSM) grew from 106 conditions in 1952 to approximately 300 in 2013, though counts vary with what is included. This review asks whether that growth reflects better recognition of real illness or systematic over-diagnosis, and what capacity the diagnostic system has to distinguish between them. The paper is a theoretical essay synthesizing sources across psychiatry, medical sociology, and the philosophy of science. It organizes the argument around three propositions: that financial conflicts of interest among DSM panel members are documented at rates that would be disqualifying elsewhere in medical standard-setting, that threshold changes cannot be validated against any external standard where no biological referent exists, and that categories alter how the people they describe present. Three cases are examined, each isolating a different mechanism: social anxiety disorder, major depressive disorder, and multiple personality disorder. The paper states which proposition each case supports and which it does not. Four further cases in which the system narrowed, refused, or reversed a category bound the argument rather than defeating it: the system is capable of contraction, and the conditions under which it contracts are empirically tractable. The second proposition is the best supported. The first is supported only in its weaker form, at the level of boundary expansion rather than category creation. Rising diagnosis rates are documented. What they indicate about underlying prevalence remains, on current evidence, underdetermined, and this paper's contribution is to specify why and to identify the reforms that would change it.

**Keywords:** Over-diagnosis; DSM; psychiatric classification; pharmaceutical industry; looping effects; philosophy of psychiatry .

## Introduction

In 2012, GlaxoSmithKline paid three billion dollars in the largest healthcare fraud settlement in United States history. The Department of Justice record documents unlawful promotion of Paxil for pediatric depression, participation in preparing a misleading journal article reporting efficacy that a clinical trial had failed to demonstrate, and the withholding of two further studies in which the drug also failed to demonstrate efficacy in patients under eighteen<sup>1</sup>. The settlement establishes that pharmaceutical misconduct reached the evidence base clinicians rely on. It does not establish that any diagnostic category was created commercially, and this paper does not use it for that purpose. It is introduced here as evidence that the informational environment surrounding psychiatric prescribing has been shaped by parties with a financial interest in its content, which is the condition under which the rest of the argument operates.

The DSM listed 106 diagnostic categories in its first edi-

tion in 1952 and approximately 300 in its fifth edition in 2013<sup>2,3</sup>. Counts vary with what is counted: diagnostic categories, subtypes, specifiers, and residual “other specified” categories are not treated consistently across editions. Figures cited in this paper count top-level diagnostic categories, comparisons across editions should be treated as indicative rather than precise, and no quantitative inference is drawn from the ratio.

This review argues that psychiatric diagnostic categories have expanded under commercial pressure in ways that clinical evidence alone does not account for, and that the diagnostic system has limited capacity to audit that expansion after the fact. The argument rests on three propositions a reader can check against evidence. First, that financial relationships between industry and DSM panel members are documented at rates that would be considered disqualifying in other areas of medical standard-setting. Second, that for categories without a biological reference standard, threshold changes cannot be validated against anything external to the committee that set them. Third, that diagnostic categories applied to people alter how those people report and interpret their symptoms, which

<sup>1</sup> The Oakridge School, Texas, USA

---

complicates later attempts to assess whether the original category was well-drawn. Each of the three cases examined here supports some of these propositions and not others, and the paper states which.

## Methods

This paper is a theoretical essay rather than a systematic review. It does not follow a reproducible search protocol and does not claim exhaustive coverage of the literature. Its method is conceptual analysis structured around three case studies selected because each isolates a different proposed mechanism: social anxiety disorder for commercial influence on category boundaries, major depressive disorder for the absence of a reference standard, and multiple personality disorder for classificatory feedback effects. The cases were selected for mechanistic clarity, not representativeness, and the paper does not claim they are typical. Sources were identified through citation-tracing from the primary theoretical literature and supplemented with targeted searches for empirical work bearing on each case. Four further cases were selected on the opposite criterion — episodes in which the diagnostic system narrowed, rejected, or reversed a category — in order to test the commercial-expansion model against evidence that would count against it. A table classifying every source by evidence type appears as Table 1. The paper presents no new data and does not estimate how many diagnoses are unwarranted.

## Source Classification

Every source used in this paper is classified in Table 1 by the kind of evidence it provides and by what it can and cannot be used to support. The classification is reported in the Methods because it is part of the method: the argument is built from unlike kinds of evidence, and the weight each source can carry differs accordingly. Of the 43 sources, 28 are peer-reviewed research articles, systematic reviews, or meta-analyses, and a further three are commentary or debate pieces published in peer-reviewed journals. The remaining twelve are primary documents that no secondary source could replace: the DSM editions under discussion, the FDA labeling and Department of Justice records quoted directly, and four scholarly monographs whose arguments the paper engages by name.

## Definitions

This paper uses over-diagnosis to mean the assignment of a diagnostic label to individuals whose presentation falls within the range of expected human response to circumstance, or whose symptoms would not have progressed to clinically significant impairment absent the label. This is distinct from four

adjacent phenomena. Misdiagnosis is the assignment of the wrong label to someone who does have a disorder. Over-treatment is the provision of intervention exceeding what a correct diagnosis warrants. Diagnostic substitution is the migration of cases between categories as definitions change, which alters category-level counts without altering the number of affected individuals. Increased case-finding is the identification of individuals who met existing criteria but had not previously been assessed, which raises diagnosis rates without constituting over-diagnosis. Rising prevalence figures are compatible with all five, and this paper's claim concerns the system's capacity to distinguish among them rather than the magnitude of any one.

## The Commercial Architecture of Diagnostic Expansion

The committees that produce the Diagnostic and Statistical Manual vote on which conditions to include, how to define them, and where to draw the boundaries between disorder and normal variation, and they do so as institutions financially entangled with the industry that profits from those boundaries expanding. A 2006 study by Cosgrove et al. examined the 170 panel members who contributed to DSM-IV and DSM-IV-TR and found that 56% had one or more financial ties to the pharmaceutical industry, with 100% of members on the Mood Disorders and Schizophrenia and Other Psychotic Disorders panels reporting such connections<sup>8</sup>. Equivalent data exist for DSM-5. Cosgrove and Krinsky compared the two editions and found that 69% of DSM-5 task force members reported industry ties, against 57% of the DSM-IV task force, and that roughly three-quarters of the 141 members serving on the thirteen DSM-5 work groups reported such ties. The concentration again fell on the panels governing the most heavily medicated categories: 100% of the Sleep/Wake Disorders panel, 83% of the Psychotic Disorders panel, and 67% of the Mood Disorders panel<sup>9</sup>. The APA introduced a formal disclosure policy between the two editions, and the rates did not fall. Transparency and independence are not the same instrument, and the DSM-5 record indicates that the first does not deliver the second. Documented financial relationships establish the conditions under which influence is possible. They do not by themselves establish that any specific criterion was changed because of them, and the distinction is maintained throughout this section.

What the wider literature establishes is that sponsorship is associated with outcomes even where no individual decision can be faulted. A Cochrane review of 75 methodological studies found industry-sponsored drug and device trials more often reported favorable efficacy results (RR 1.27, 95% CI 1.17–1.37) and more often reached favorable conclusions (RR 1.34, 95% CI 1.19–1.51) than trials funded otherwise, and that the difference was not explained by standard risk-of-bias assess-

**Table 1** Classification of all sources by evidence type, with what each source can and cannot be used to support.

| Source                                           | Evidence type           | Supports / does not support                                                                             |
|--------------------------------------------------|-------------------------|---------------------------------------------------------------------------------------------------------|
| DOJ, GSK settlement (2012) <sup>1</sup>          | Legal record            | Misconduct and suppressed trial data. NOT category creation.                                            |
| DSM-I (1952) <sup>2</sup>                        | Primary document        | 1952 category counts. Conventions differ across editions.                                               |
| DSM-III (1980) <sup>4</sup>                      | Primary document        | Entry of social phobia and MPD into the manual.                                                         |
| DSM-III-R (1987) <sup>5</sup>                    | Primary document        | Pre-1994 MPD criteria, including 'full control'.                                                        |
| DSM-IV (1994) <sup>6</sup>                       | Primary document        | Renamed DID criteria and added amnesia requirement; establishes non-comparability.                      |
| DSM-5 (2013) <sup>3</sup>                        | Primary document        | Category counts; bereavement exclusion removal; Section III placement.                                  |
| DSM-5-TR (2022) <sup>7</sup>                     | Primary document        | That the categorical architecture was retained at text revision.                                        |
| Cosgrove et al. (2006) <sup>8</sup>              | Empirical               | Ties on DSM-IV panels. NOT other editions, NOT causation.                                               |
| Cosgrove & Krinsky (2012) <sup>9</sup>           | Empirical               | DSM-5 panel ties; that disclosure did not reduce them. NOT that any criterion changed.                  |
| Lundh et al. (2017) <sup>10</sup>                | Systematic review       | Sponsorship and favorable results. Trials, NOT committees; used as analogy.                             |
| Liebowitz et al. (1985) <sup>11</sup>            | Review                  | That social phobia was recognized and called neglected before the campaign.                             |
| Lane (2008) <sup>12</sup>                        | Historical              | The Cohn & Wolfe campaign. Secondary, not a marketing record.                                           |
| FDA labeling, NDA 20-031/S-023 <sup>13</sup>     | Regulatory document     | Exact wording of the SAD indication and shyness sentence. NOT how it is applied.                        |
| Healy (2001) <sup>14</sup>                       | Debate piece            | Marketing expanded recognition, NOT created the category. Partly against the strong thesis.             |
| Sheehan (2001) <sup>15</sup>                     | Debate piece            | The opposing position. Counterweight, not support.                                                      |
| Drescher (2015) <sup>16</sup>                    | Historical              | 1973–74 removal of homosexuality. Disconfirming case.                                                   |
| Carpenter & van Os (2011) <sup>17</sup>          | Clinical review         | The case FOR attenuated psychosis syndrome; shows what was declined.                                    |
| Kulage et al. (2014) <sup>18</sup>               | Meta-analysis           | DSM-5 autism criteria reduced the diagnosed population. Tightened threshold.                            |
| Loomes et al. (2017) <sup>19</sup>               | Meta-analysis           | Under-ascertainment of autism in girls. Expansion as correction.                                        |
| Regier et al. (2013), II <sup>20</sup>           | Empirical               | Test-retest reliability by category. Reliability only, NOT validity.                                    |
| Regier et al. (2013), I <sup>21</sup>            | Methods                 | Field trial design; what the kappa figures mean.                                                        |
| Horwitz & Wakefield (2007) <sup>22</sup>         | Argument                | Critique of DSM-IV MDD. Predates DSM-5.                                                                 |
| Wakefield & First (2012) <sup>23</sup>           | Empirical review        | Case against removing the bereavement exclusion. One side of a live dispute.                            |
| Zachar, First & Kendler (2017) <sup>24</sup>     | Participant account     | How the bereavement decision was reached. Partly by participants.                                       |
| Levis et al. (2020) <sup>25</sup>                | IPD meta-analysis       | PHQ-9 overstates prevalence vs interview. Screening only.                                               |
| Hacking, looping effects (1995) <sup>26</sup>    | Argument                | The mechanism. Cannot support a prevalence conclusion.                                                  |
| Hacking, Rewriting the Soul (1995) <sup>27</sup> | Historical              | MPD case estimates. Estimates, not a census.                                                            |
| Paris (2012) <sup>28</sup>                       | Peer-reviewed review    | Independent account of the DID rise and decline.                                                        |
| GBD 2019 Collaborators (2022) <sup>29</sup>      | Burden estimate         | Ranking and burden of depressive disorders.                                                             |
| Cipriani et al. (2018) <sup>30</sup>             | Network meta-analysis   | Antidepressant efficacy vs placebo. NOT diagnostic validity.                                            |
| Henderson & Thornicroft (2009) <sup>31</sup>     | Programme description   | Campaign aims at launch. CANNOT report outcomes.                                                        |
| Evans-Lacko et al. (2013) <sup>32</sup>          | Empirical               | Actual Time to Change outcomes, including null results.                                                 |
| Kvaale et al. (2013) <sup>33</sup>               | Meta-analysis           | Biogenetic explanations reduce blame but raise pessimism. Harm of labelling.                            |
| Layton et al. (2018) <sup>34</sup>               | Cohort                  | Relative-age effect on ADHD diagnosis. Specific to ADHD.                                                |
| Ruscio (2010) <sup>35</sup>                      | Taxometric + predictive | SAD dimensional at the margin; dimensional diagnosis outpredicts categorical. NOT the shyness boundary. |
| Kollman et al. (2006) <sup>36</sup>              | Taxometric              | Prior dimensional finding in outpatients. Social fear, not full criteria.                               |
| Davidson et al. (1994) <sup>37</sup>             | Empirical               | Prevalence shifts with the impairment threshold.                                                        |
| Insel et al. (2010) <sup>38</sup>                | Programme statement     | RDoC rationale. A research framework, NOT a nosology.                                                   |
| Kotov et al. (2017) <sup>39</sup>                | Structural model        | HiTOP as an empirical alternative. Contested.                                                           |
| Borsboom & Cramer (2013) <sup>40</sup>           | Methodological review   | Network approaches. Theoretical alternative.                                                            |
| Caspi et al. (2014) <sup>41</sup>                | Cohort                  | The <i>p</i> factor. Comorbidity structure, NOT any specific boundary.                                  |
| First et al. (2021) <sup>42</sup>                | Comparative analysis    | ICD-11 / DSM-5 divergence across 103 disorders.                                                         |
| Batstra & Frances (2012) <sup>43</sup>           | Proposal                | Stepped diagnosis. A proposal, not an evaluated intervention.                                           |

---

ment<sup>10</sup>. That concerns trials rather than committees and is offered as analogy, not direct evidence about the DSM. Its relevance is that a systematic tilt surviving procedural safeguards, traceable to no single decision, is demonstrably real somewhere in medicine rather than merely conceivable.

Social phobia entered the DSM in its third edition in 1980, and its criteria were revised in DSM-III-R and again in DSM-IV, well before any pharmaceutical marketing campaign targeted it<sup>4-6</sup>. The category was clinically recognized but received comparatively little research and treatment attention, a state of affairs the psychiatric literature of the period named directly: Liebowitz and colleagues titled their 1985 review in the Archives of General Psychiatry “Social phobia: review of a neglected anxiety disorder,” and reported major uncertainties in classification, prevalence, severity, etiology, assessment, and treatment<sup>11</sup>. The phrase is the literature’s own, and it is dated fourteen years before the marketing campaign at issue. What changed in the late 1990s was not the existence of the category but the commercial attention paid to it. SmithKline Beecham, facing patent expiry on Paxil, pursued an additional indication. The company engaged the public relations firm Cohn & Wolfe, which built an awareness campaign around the phrase “imagine being allergic to people,” and funded the clinical trials supporting the application<sup>12</sup>. The FDA approved Paxil for social anxiety disorder in 1999. The approved labeling defines the indication by reference to the diagnostic manual and then draws the boundary itself. It states that Paxil “is indicated for the treatment of social anxiety disorder, also known as social phobia, as defined in DSM-IV (300.23),” and closes the indication with the sentence “Lesser degrees of performance anxiety or shyness generally do not require psychopharmacological treatment”<sup>13</sup>. That last sentence is the boundary at issue here, and it is worth noting that the regulatory document draws it explicitly. The label anticipates the objection; what it cannot do is establish where the line falls, or prevent the marketing of the indication from moving it.

The effect of the campaign is contested, and the contest matters because the strongest version of the commercial-construction thesis does not survive it. In 2001 the Western Journal of Medicine ran a paired debate on whether marketing had manufactured social anxiety disorder. Healy, arguing the critical side, did not claim the category was created commercially: he described social phobia as “relatively unrecognized in the West,” called the renaming “symbolic of what can happen” when a condition becomes a marketing vehicle, and located the harm in suppressed trial data and diverted research<sup>14</sup>. Sheehan, opposing, held that the disorder had “long been trivialized and even ignored” and that industry funded the trials at the urging of academic researchers<sup>15</sup>. They disagree about the value of what the campaign did and agree about what it did not do: neither claims the category was

commercially invented. This paper adopts the position both sides share. What the record supports is that marketing broadened public recognition, increased case-finding, encouraged pharmacological treatment, and blurred the boundary between disabling social anxiety and ordinary shyness. It does not support the claim that commercial interest created the diagnostic category. The distinction matters for the argument of this paper: a category can be legitimately established and subsequently expanded past the evidence supporting its boundaries, and boundary expansion is the mechanism at issue here.

The growth of the DSM across precisely the decades in which pharmaceutical companies became the dominant funders of psychiatric research is consistent with a system responding to its incentive structure. Consistency is not proof, and competing explanations exist: increased research activity, the professionalization of diagnostic specificity, and pressure from insurers for codable categories all predict expansion independently of commercial influence. A further class of evidence bears more directly against the model, and the next section sets it out.

### Cases the Argument Does Not Explain

A model that accounts only for expansion is not a model of the system; it is a description of one of its behaviors. If commercial incentive were the dominant force acting on diagnostic boundaries, the DSM should expand monotonically, and categories should not narrow, reverse, or fail to be adopted. Four cases show otherwise, and none of them is well explained by the account given above.

The removal of homosexuality is the clearest. The APA Board of Trustees voted in December 1973 to strike homosexuality from DSM-II, replacing it with the compromise category “sexual orientation disturbance”; the vote was challenged and removal confirmed by membership referendum in 1974, the residual category finally eliminated in DSM-III-R in 1987<sup>16</sup>. No pharmaceutical interest stood to gain, and none is recorded as having pressed for it. The change was driven by activist pressure, accumulating evidence that the classification lacked validity, and professional argument conducted largely in public. The case concedes three things: a category can be removed outright, the process can be moved by evidence and political pressure rather than commerce, and the deliberation can be visible enough to reconstruct afterward — which is what the archival opacity discussed later is said to prevent.

The second case is a proposal the committee declined. Attenuated psychosis syndrome was advanced for DSM-5 with an early-intervention rationale, its proponents arguing that the evidence for predictive validity sufficed to establish the diagnostic class<sup>17</sup>. A category identifying sub-threshold psychotic symptoms in young people would have created a large treatment-eligible population, and the commercial-expansion

---

model predicts adoption. The task force instead placed the condition in Section III, the section reserved for conditions requiring further study, rather than adopting it as a full diagnostic category<sup>3</sup>. The disposition is not a clean rejection — the presentation remains codable under other specified schizophrenia spectrum disorder — and should be cited with that qualification. What it shows is that the process can withhold categorical status from a proposal that would have expanded the diagnosed population.

The third case is a threshold that was tightened. DSM-5 folded Asperger's disorder and pervasive developmental disorder not otherwise specified into a single autism spectrum disorder category with more restrictive criteria. A systematic review and meta-analysis of fourteen studies found that the change reduced the proportion of previously diagnosed individuals who continued to meet criteria, with reductions across studies ranging from 7.3% to 68.4%; the effect fell hardest on those previously diagnosed with Asperger's disorder and PDD-NOS<sup>18</sup>. This is contraction and consolidation rather than expansion, and it also illustrates diagnostic substitution as defined earlier: category-level counts moved without any change in the number of affected individuals.

The fourth case runs in the opposite direction and is the most useful to the argument, because it shows expansion functioning as correction. A meta-analysis of 54 prevalence studies covering 13.8 million participants found a pooled male-to-female odds ratio for autism of 4.20 (95% CI 3.84–4.60), but 3.32 (95% CI 2.88–3.84) among high-quality studies and 3.25 (95% CI 2.93–3.62) among studies that screened the general population rather than recruiting people who already held a diagnosis — against 4.56 (95% CI 4.10–5.07) in studies restricted to those with an existing diagnosis<sup>19</sup>. The authors conclude that the true ratio is closer to three to one than the conventional four to one, and that girls meeting criteria are at disproportionate risk of not being diagnosed. Wider recognition here does not manufacture cases; it finds cases the narrower criteria were missing. Any account of diagnostic expansion that cannot accommodate this is incomplete.

These cases do not defeat the argument of this paper, and it would be a misreading to take them as offsetting entries in a ledger. They bound it. The diagnostic system is capable of contraction, refusal, and self-correction, and the conditions under which it exercises those capacities appear to include organized external pressure, evidence of poor reliability, and the absence of a treatment market attached to the category in question. That is a testable claim, and it converts the thesis from an assertion that the system expands into the more defensible and more interesting question of when it does. The paper's remaining claim is narrower than the one it started with: not that expansion is the system's only behavior, but that the system's capacity to audit expansion is weaker than its capacity to produce it.

## The Absence of a Biological Reference Standard

Psychiatric diagnosis is not alone in medicine in lacking definitive biomarkers. Migraine, irritable bowel syndrome, fibromyalgia, and many rheumatological conditions are diagnosed syndromically, and numerous conditions with established biological correlates still rely on probabilistic thresholds set by consensus. Psychiatry also possesses substantial biological research: neuroimaging findings, genetic association studies, physiological measures, and validated clinical instruments with documented reliability. What distinguishes psychiatric classification is not the total absence of biological evidence but the absence of any measure with sufficient diagnostic specificity to constrain where a committee places a threshold. In conditions where a biomarker exists, a proposed criterion change can be evaluated against rates of confirmed disease. For most DSM categories, no such external check is available, and threshold decisions are validated primarily against clinical judgment and field-trial reliability rather than against an independent standard of what the disorder is.

The DSM-5 field trials indicate how much weight that substitute can bear. Regier and colleagues measured test-retest agreement between independent clinicians assessing the same patient on separate occasions across a large multi-site sample<sup>20,21</sup>. Results were mixed rather than uniformly poor: five diagnoses reached very good agreement and nine were good. But four fell into the range the investigators themselves called questionable, an intraclass kappa between 0.20 and 0.39, and these included the two highest-prevalence categories in the manual. Major depressive disorder reached 0.28. One proposed category, mixed anxiety-depressive disorder, fell below 0.20 and was not adopted<sup>20</sup>. The point is structural rather than about the absolute figures, which are themselves contested. Reliability measures agreement between clinicians applying a criterion set; it cannot measure whether the set is drawn in the right place, because both clinicians read from the same manual. A category can be perfectly reliable and still be drawn around the wrong population.

Horwitz and Wakefield's analysis addresses the DSM-IV framework, in which major depressive disorder was defined by five or more symptoms persisting two weeks or longer, with a narrow exclusion for recent bereavement<sup>22</sup>. Their argument is that the criteria captured proportionate responses to loss and adversity alongside disorder, and that the bereavement exclusion was too narrow to correct for this. DSM-5 removed the bereavement exclusion entirely in 2013<sup>3</sup>. The APA's stated rationale was that bereavement should not protect an individual from a depression diagnosis when criteria are otherwise met, that depression following loss responds to treatment, and that the exclusion risked withholding care. Critics argued the removal would pathologize normal grief.

The exchange that followed is an unusually well-

---

documented instance of a threshold decision contested on empirical grounds. Wakefield and First argued that the reviews cited in support of removal had not tested the exclusion: they compared standard depression against mixed samples of bereavement-related depression, most of which would not have satisfied the exclusion's own conditions of duration under two months and absence of specific severe symptoms. They further cited National Comorbidity Survey data in which individuals with excluded bereavement-related depression showed three-year recurrence of 4.3%, close to those with no depression history and well below the 14.7–27.2% in standard major depression, and concluded the evidence supported retaining the exclusion<sup>23</sup>. Zachar, First and Kendler later reconstructed the decision from the inside and reached a more equivocal verdict: the removal turned as much on how the burden of proof was allocated as on any decisive finding<sup>24</sup>. The empirical literature addresses duration, severity, functional impairment, and whether contextual assessment can reliably distinguish the two, and on none of these has it converged.

The point this paper takes from that dispute is limited. Whether the DSM-5 change increased false positives is an open empirical question, and its being open is itself the issue: the criteria contain no mechanism for settling it internally, because there is no reference standard against which a bereaved person meeting symptom criteria can be classified as ill or not ill independent of the criteria themselves. The Zachar, First and Kendler reconstruction sharpens this rather than softening it. When the participants in a decision, writing afterward with access to the deliberative record, cannot identify the finding that settled it, the absence of an external standard is not a philosophical observation about psychiatry but a practical description of how the decision was actually made.

A second infrastructure sustains the pattern. Diagnosis has become a primary language for distress: access to insurance reimbursement, workplace accommodation, and educational provision is frequently gated behind a diagnostic code, which creates demand for diagnosis independent of clinical need. This paper does not claim that no constraints operate. Functional-impairment requirements, differential diagnosis, structured assessment instruments, professional guidelines, and payer criteria all restrict diagnosis in practice. The claim is narrower: these constraints operate on the application of criteria to individuals, not on where the criteria are set, and none of them provides a mechanism for evaluating whether a threshold was correctly placed.

Two bodies of evidence show specific constraints failing as the claim predicts. Screening instruments are often treated as an external check on clinical impression. An individual-participant-data meta-analysis of 44 studies covering 9,242 participants found that 24.6% scored at or above the standard PHQ-9 cutoff of 10, against a structured-interview prevalence of major depression of 12.1% in the same samples<sup>25</sup>.

The instrument is not defective; screening is designed to over-include. It is defective as a check on where the threshold sits, which is how such figures are often used. The second concerns applying criteria to individuals, where constraints should work best. In a cohort of 407,846 US children, those born in August — the youngest in their cohort where a September 1 school-entry cutoff applied — were diagnosed with ADHD at 85.1 per 10,000 against 63.6 for September-born children, a difference of 21.5 per 10,000 (95% CI 8.8–34.0), with no comparable effect in states without the cutoff<sup>34</sup>. Relative age is not a feature of any diagnostic criterion. That it predicts diagnosis at this magnitude indicates criterion-level constraints did not bind against a variable they do not mention.

### The Looping Effect

A beetle does not change its behavior simply because a scientist has named it. Humans, however, are not indifferent to classification in the way that beetles are. Ian Hacking, in his 1995 paper “The Looping Effects of Human Kinds,” argued that diagnostic categories actively transform the people they describe<sup>26</sup>. When a person receives a diagnosis, they read about it, reinterpret their past through it, and begin presenting symptoms in the ways the category predicts. The category reshapes the person, and as people change to fit the category, the category updates to reflect them, producing a feedback loop between clinical classification and human identity.

This mechanism requires careful scoping. That a diagnostic category shapes how symptoms are reported and interpreted does not establish that the underlying condition is absent or fabricated. A person with a well-founded diagnosis may still learn the category's vocabulary, reinterpret their history through it, and present in ways the category anticipates. Looping affects the expression and reporting of a condition, and the inference from altered expression to constructed disorder does not follow. What looping does establish is narrower: that patient presentation cannot serve as independent confirmation of a category's validity, because presentation is partly a product of the category. This removes one line of evidence that might otherwise validate a boundary. It does not by itself show any boundary to be wrongly drawn.

The condition discussed below was listed as multiple personality disorder in DSM-III (1980) and DSM-III-R (1987), and renamed dissociative identity disorder in DSM-IV (1994)<sup>4–6</sup>. This paper uses the term contemporaneous with the period under discussion.

Hacking reports that as of 1972 multiple personality appeared to be a curiosity, with fewer than a dozen cases reported in the preceding fifty years, and that by 1986 some six thousand patients were thought to have been diagnosed<sup>27</sup>. These are contemporaneous estimates circulating in the clinical literature rather than a census, and Hacking presents them as

---

such; the argument here uses the direction and steepness of the change, not the precise counts. The trajectory is independently described in the peer-reviewed literature, where Paris characterises the rise and subsequent decline of the diagnosis as following the course of a clinical fashion<sup>28</sup>. Clinicians became skeptical, the attention once given to this condition receded, and it was renamed to dissociative identity disorder in 1994. Reported diagnosis rates declined substantially thereafter.

Several explanations compete for this pattern, and the case does not adjudicate among them. Increased clinical awareness following DSM-III inclusion would predict a rise in identification of cases previously missed. Shifts in referral patterns concentrated cases among a small number of specialist clinicians, which affects observed prevalence independent of true prevalence. Media attention, including widely circulated case narratives, plausibly influenced both patient presentation and clinician expectation. The recovered-memory controversy of the early 1990s and the litigation that followed altered the professional risk of making the diagnosis, which would depress rates regardless of the condition's status. The renaming in DSM-IV accompanied criteria revisions, making pre- and post-1994 counts non-comparable. The revision is specific enough to state. DSM-III-R had required two or more distinct personalities, at least two of which "recurrently take full control of the person's behavior"<sup>5</sup>. DSM-IV renamed the condition, replaced "personalities" with "identities or personality states," dropped "full" from the control requirement, and added a criterion absent from the earlier definition: inability to recall important personal information beyond ordinary forgetfulness, plus an exclusion for substance-induced and medical causes<sup>6</sup>. The two amendments pull in opposite directions — dropping "full" loosens, requiring amnesia tightens — so counts taken before and after 1994 are not measuring the same construct. The trajectory is consistent with a category whose observed prevalence tracked clinical and cultural attention rather than an underlying rate. It is also consistent with a real but rare condition that was first under-recognized, then over-diagnosed by a subset of practitioners, then avoided. Distinguishing these requires evidence the case record does not contain, which is the point at issue.

### Weighing the Case for Diagnostic Expansion

Diagnostic expansion has nonetheless produced consequences that complicate the argument made so far. For most of the twentieth century, depression went untreated and unrecognized. The burden is substantial on current estimates: the Global Burden of Disease 2019 analysis of twelve mental disorders across 204 countries ranked depressive disorders thirteenth among all causes of disability-adjusted life-years worldwide, with depressive disorders accounting for the

largest single share of mental-disorder DALYs, and found no evidence of a global reduction in mental-disorder burden since 1990<sup>29</sup>. Undertreatment is documented and ongoing, not merely historical. The expansion gave millions of people a language for experiences that had previously been dismissed, and with that language came access to treatment, legal accommodation, and social legitimacy.

The same commercial infrastructure criticized above also produced lithium stabilization protocols for bipolar disorder, clozapine for treatment-resistant schizophrenia, and SSRIs whose efficacy in major depression is documented across a large network meta-analysis. Cipriani and colleagues pooled 522 double-blind trials enrolling 116,477 participants and found all 21 antidepressants examined more efficacious than placebo, with odds ratios for response ranging from 2.13 for amitriptyline to 1.38 for reboxetine<sup>30</sup>. The effect is real and the interval between the highest and lowest estimate is wide; this paper draws no conclusion from the ordering of individual agents.

Henderson and Thornicroft described the design and aims of England's Time to Change campaign at its launch, and that paper cannot report outcomes<sup>31</sup>. The outcome evaluation, covering 2009 to 2012, found a statistically significant improvement in intended behaviour of 0.07 standard deviation units (95% CI 0.01–0.14) and a non-significant trend toward improved attitudes ( $p = 0.08$ ), but no significant improvement in either knowledge or reported behaviour<sup>32</sup>. That is a modest and partial result rather than a broad reduction in stigma. It is reported in full because the direction of the finding runs against this paper's own convenience: a weaker benefit from public mental-health campaigning makes the case for expansion less impressive, and the figure should not be inflated on either side of the argument.

Harms of expansion are also documented and belong in the same weighing: adverse effects of unnecessary treatment, the stigma that a diagnosis can itself confer, diagnostic instability over time, and unequal access to diagnosis across populations. The stigma harm is the best evidenced and the least intuitive. A meta-analytic review of experimental and correlational studies found that biogenetic explanations of mental disorder reduce attributions of blame, which is the effect such explanations are usually advocated for, while simultaneously increasing perceived dangerousness and prognostic pessimism — the expectation, held by the public, by clinicians, and by diagnosed people themselves, that the condition will not improve<sup>33</sup>. The medicalizing move that relieves moral judgment also imports a prognosis. The relative-age findings discussed earlier bear on unequal application<sup>34</sup>, and the screening literature on over-inclusion<sup>25</sup>. The system that compromised psychiatric diagnosis is also the system that made its expansion possible, and that tension runs through every piece of evidence examined here.

---

## The Limits of That Case

Real psychiatric conditions exist, the DSM identified many of them correctly, and the treatments built around them have extended and improved lives. What that case does not establish is that the system possesses a reliable mechanism for evaluating, after the fact, whether a given category's boundaries were set by evidence or by pressure.

Lithium and clozapine entered psychiatric practice through the same institutional apparatus that produced the expansion documented above. Shared institutional origin, however, does not prevent categories from being distinguished by other means, and the earlier version of this argument overstated the case. Diagnostic validity can be assessed through independent replication across research groups, diagnostic reliability testing, predictive validity against longitudinal outcomes, convergent evidence from multiple assessment modalities, and treatment-response patterns that differentiate the category from its neighbors. The claim that categories differ on these measures is load-bearing, and the support for it is partial. The DSM-5 field trials provide the most direct evidence: the categories treated here as well-established, including the psychotic and bipolar disorders, reached higher test-retest agreement than major depressive disorder and generalized anxiety disorder<sup>20</sup>. That runs in the predicted direction. It is also weaker than the claim requires, for the reason given earlier — reliability is agreement about a criterion, not evidence the criterion sits in the right place — and the trials did not assess social anxiety disorder, the category the commercial argument concerns. The honest position is therefore: differential reliability is documented, differential predictive and construct validity is plausible and partly supported by the alternative-framework literature below, and a direct comparison of the categories identified here as commercially expanded against those treated as well-founded has not been carried out. That comparison is the study this argument most needs and does not have.

One piece of it does exist, and it bears directly on social anxiety disorder. Ruscio submitted indicators representing the DSM-IV criteria for social anxiety disorder to taxometric analysis in a subsample of 2,166 adults from the National Comorbidity Survey Replication who reported excessive social fear. Multiple procedures and consistency tests converged on a dimensional rather than a categorical solution<sup>35</sup>, replicating an earlier taxometric study of social fear in more than two thousand anxiety and mood disorder outpatients<sup>36</sup>. Measured prevalence for the disorder is correspondingly sensitive to where the criteria are set, and particularly to the level of impairment required<sup>37</sup>.

The structural finding is not the most useful part. Ruscio compared the DSM-IV categorical diagnosis against a dimensional diagnosis built from the same interview items, using

discrete-time survival models to predict outcomes arising after the onset of social anxiety. The dimensional diagnosis predicted better on ten of eleven outcomes and significantly better on six, with large differences: subsequent mood disorder ( $d = 0.95$ ), suicidal ideation ( $d = 1.86$ ), suicide attempt ( $d = 4.71$ ), psychotherapy ( $d = 0.89$ ), psychotropic medication ( $d = 2.17$ ), and any professional mental health treatment ( $d = 0.86$ )<sup>35</sup>. Because both diagnoses drew on the same items, the comparison isolates the threshold itself rather than the amount of information collected. What the threshold does is discard clinically significant information about who will subsequently become suicidal and who will seek treatment.

Two qualifications are required, and the second constrains the claim more. Dimensional structure does not establish that the category is invalid, and Ruscio does not argue that it does. More importantly, the analysed sample was restricted to respondents already reporting excessive social fear with distress or avoidance, roughly half of whom met criteria for lifetime social anxiety disorder. The analysis therefore tests the boundary between diagnosed and subthreshold cases among people who are already anxious, not the boundary between social anxiety and ordinary shyness, and Ruscio states plainly that a taxonic boundary at milder levels remains possible. That limitation cuts against the broadest version of this paper's argument and is reported for that reason. What survives is narrower and still substantial: at the margin where over-diagnosis would occur, there is no natural joint at which the line must be drawn, and the line as drawn measurably costs prognostic information. The threshold's location is a decision rather than a discovery, in the one category whose commercial history is best documented.

A further distinction is required. Treatment efficacy establishes clinical utility, not diagnostic validity. That a drug helps patients selected by a set of criteria shows the criteria identify a group who benefit; it does not establish that the criteria carve the condition at its boundaries, and it says nothing about patients near the threshold who were included as the threshold moved. Conversely, a category with contested validity can retain clinical utility if it reliably identifies people who benefit from available intervention. The argument in this paper concerns validity at the boundaries of categories, not the utility of treatments delivered under them.

Once a category enters the DSM, whether through scientific discovery or commercial lobbying, it begins reshaping the people it describes. Their reinterpreted histories, altered presentations, and transformed identities become part of the evidentiary basis for the category's continued existence. This weakens retrospective evaluation without foreclosing it: the validity measures listed above remain available, and they do not depend on patient self-report alone.

---

## Alternative Frameworks and What They Reveal

The categorical model criticized here is not the only one available, and the existence of developed alternatives is itself evidence about the DSM's boundaries. The Research Domain Criteria framework, launched by the National Institute of Mental Health in 2010, abandoned DSM categories as the unit of research analysis in favour of dimensional constructs spanning genetic, circuit, physiological and behavioural levels; its authors were explicit that it was a research framework rather than a replacement nosology<sup>38</sup>. The Hierarchical Taxonomy of Psychopathology derives its structure from the observed covariation of symptoms rather than committee consensus, organizing psychopathology into spectra that cut across DSM boundaries<sup>39</sup>. Network approaches treat disorders not as latent entities generating symptoms but as systems of symptoms that cause one another<sup>40</sup>. And factor-analytic work on cohort data has repeatedly recovered a single general dimension of psychopathology, the *p* factor, that accounts for psychiatric comorbidity better than the correlated-categories model<sup>41</sup>.

These programmes disagree with one another substantially, and none has displaced the DSM in clinical practice. Their evidential value for this paper does not depend on any of them being right, and the claim drawn from them should be stated carefully. RDoC's stated objection was that DSM categories are not useful units for research, which is not the same as the claim that they are misplaced. What the four programmes share is weaker and still informative: each concluded that the categorical boundaries do not correspond to the structure its own methods recover, and each built an alternative unit of analysis in response. A boundary that survives because it is entrenched in reimbursement codes, regulatory labels and research infrastructure, rather than because the evidence locates it there, is a boundary of the kind this paper has been describing. The DSM-5 text revision published in 2022 retained the categorical architecture unchanged in this respect<sup>7</sup>, so the alternatives remain alternatives.

The point is reinforced by the comparison between the two operative classification systems. First and colleagues compared ICD-11 and DSM-5 requirements for 103 corresponding disorders and rated 20 of them (19.4%) as differing in major respects, with a further 42 (40.8%) differing in minor definitional respects and only 31 (30.1%) essentially identical<sup>42</sup>. Two committees, working from substantially the same literature over substantially the same period, produced materially different criteria for roughly a fifth of the disorders they both recognize. That divergence is direct evidence bearing on this paper's second proposition. Where an external reference standard exists, two expert bodies reading the same evidence converge. Where none exists, the placement of the boundary reflects the priorities of the body that placed it, and the ICD-DSM comparison shows how much room those priorities

occupy.

## Implications for Reform

The preceding argument identifies an asymmetry rather than a conspiracy: the mechanisms that expand diagnostic categories are stronger than the mechanisms that audit them. That framing determines what reform would have to consist of. It is not a matter of removing commercial influence from psychiatry, which is neither achievable nor obviously desirable given that the same infrastructure produced the treatments discussed above. It is a matter of strengthening the audit side until it is comparable in force to the expansion side. Four measures follow from the specific weaknesses this paper has identified, and each is directed at a failure the paper has evidenced rather than at the general problem.

The first concerns who sets thresholds. The DSM-5 experience shows disclosure alone does not reduce financial entanglement: the APA introduced a disclosure requirement between editions and the proportion of panel members reporting industry ties rose rather than fell<sup>9</sup>. The operative rule should therefore govern composition, not merely reporting. Panels setting or revising thresholds should include substantial membership with no industry relationships during the revision period and for a defined interval before it, with the strictest requirement on panels governing categories with the largest attached treatment markets — which are, on the published data, where ties are currently most concentrated. A weaker but immediately available version is to require that any threshold change be published with the evidence considered and the alternatives rejected, so the reasoning survives for later evaluation<sup>24</sup>.

The second concerns what thresholds are checked against. Where no biological reference standard exists, the alternative is not to invent one but to require that threshold changes carry a prospective commitment to measurement. A revision expected to change the diagnosed population should be accompanied at adoption by a pre-registered plan specifying what would count as evidence that it had produced false positives, and by a scheduled audit at a defined interval. The bereavement exclusion is the case in point: the removal was contested on empirical grounds before it was made, and no mechanism existed to gather the deciding evidence afterward<sup>23,24</sup>. Scheduled audits would not resolve such disputes in advance; they would convert them from permanent disagreements into questions with a date attached. Routine collection of dimensional alongside categorical measures serves the same end, since dimensional data allow a threshold to be moved analytically after the fact<sup>39,41</sup>.

The third concerns evaluation after adoption, the fourth restraint at diagnosis. Post-marketing evaluation of a diagnostic change — on prevalence, prescribing, and outcomes in

the population it adds — should be funded independently of any party with an interest in the answer, on the reasoning the Cochrane sponsorship review supports: the association between sponsorship and favorable findings survives standard bias controls<sup>10</sup>. Clinically, stepped diagnosis answers threshold uncertainty without waiting for nosological reform. Batstra and Frances propose that in milder and more ambiguous presentations clinicians proceed through watchful waiting and staged assessment rather than diagnosing at first contact<sup>43</sup>. The relative-age findings show where such restraint would bind: a diagnosis predicted by a child’s birth month is made where developmental variation and disorder are hardest to distinguish<sup>34</sup>. None of this addresses the deepest problem raised here, that a category in circulation is already reshaping the people it describes. These measures aim at the more tractable objective of making the next threshold change easier to evaluate than the last.

Discussion

The practical weight of the argument lands on a single figure: the patient. A person diagnosed within a commercially expanded category and a person diagnosed with a well-established condition are processed by the same criteria, with no institutional mechanism available to either of them for knowing which category they occupy. That asymmetry of information is structural, and looping effects compound it over time.

Rising diagnosis rates admit more than one explanation, and the evidence assembled here does not settle which dominates. Epidemiological studies, diagnostic validity research, longitudinal cohorts, and treatment-response data all bear on the question, and the argument in this paper does not displace them. What the three cases establish is narrower: that the mechanisms available for auditing diagnostic expansion are weaker than the mechanisms driving it, and that the asymmetry is structural rather than accidental. That is a claim about difficulty, not impossibility. It suggests where the burden of proof should sit when a category expands, and it identifies which features of the current system make that burden hard to discharge. The disconfirming cases qualify this in a way worth stating plainly in closing: the system has contracted, refused, and corrected itself, and it did so under conditions that can be described. The question this paper leaves open is not whether audit is possible but what makes it happen.

Because the three cases are asked to do different work, it is worth setting out plainly which of them supports which proposition, and which do not (Table 2). No case supports all three, and each proposition rests on a different case.

The pattern is the point. Proposition 1 rests on one case and is supported there only in its weaker form. Proposition 2 is the best supported and draws on all three. Proposition

Table 2 Which of the three propositions each case supports, and which it does not.

| Case                                                           | P1: commercial influence on boundaries                                                                                       | P2: no external reference standard                                                                           | P3: looping effects                                                                        |
|----------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Social anxiety disorder                                        | Supports at the level of boundary expansion and treatment uptake. NOT commercial creation of the category <sup>14,15</sup> . | Partly. No natural joint at the threshold; prevalence moves with the impairment criterion <sup>35,37</sup> . | Not examined for this case.                                                                |
| Major depressive disorder (bereavement exclusion)              | No. No commercial actor appears in the record of the decision.                                                               | Directly. The removal was contested empirically and could not be settled internally <sup>23,24</sup> .       | Not examined for this case.                                                                |
| Multiple personality / dissociative identity disorder          | No. No treatment marker was attached to the category.                                                                        | Partly. The 1994 criteria change makes pre- and post-counts non-comparable <sup>5,6</sup> .                  | Yes. The clearest instance here of classification altering presentation <sup>26-28</sup> . |
| Disconfirming cases (removal, refusal, tightening, correction) | Counts against. The system contracted and self-corrected without commercial pressure <sup>16-19</sup> .                      | Neutral.                                                                                                     | Neutral.                                                                                   |

3 rests on a single case that the other two do not speak to. A reader who accepts the paper’s second proposition is not thereby committed to its first or third.

This review has limits of its own. The method here is conceptual. It leans on a handful of cases instead of a survey of every diagnostic category, and it puts no number on how many diagnoses are unwarranted. The three cases were selected because each isolates a mechanism cleanly, and cases selected on that basis are not a random sample of diagnostic history. The archival record required to trace individual criterion changes to the financial relationships of specific panel members is not fully public, which limits how far causal claims can be pressed and is itself a constraint on external evaluation of the process. That limitation was tested rather than assumed. An attempt was made to trace a single criterion change end to end — proposal, proposer, financial relationships, deliberation, evidence considered, and subsequent effect on prevalence and prescribing — and it did not succeed for any change examined here. The bereavement exclusion came closest, because participants published a retrospective account<sup>24</sup>, and even there the record does not connect individual panel members’ financial relationships to positions taken. The DSM-5 conflict-of-interest data are reported at panel level rather than linked to votes<sup>9</sup>, and votes are not published. The failure is reported because it is itself evidence for the paper’s claim about auditability.

References

1 United States Department of Justice. GlaxoSmithKline to plead guilty and pay \$3 billion to resolve fraud allegations and failure to report safety data. 2012. <https://www.justice.gov/archives/opa/pr/glaxo>

- osmithkline-plead-guilty-and-pay-3-billion-resolve-fraud-allegations-and-failure-report.
- 2 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 1st ed. American Psychiatric Association, 1952.
- 3 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 5th ed. American Psychiatric Association, 2013.
- 4 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 3rd ed. American Psychiatric Association, 1980.
- 5 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 3rd ed., revised. American Psychiatric Association, 1987.
- 6 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 4th ed. American Psychiatric Association, 1994.
- 7 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 5th ed., text revision. American Psychiatric Association, 2022.
- 8 L. Cosgrove, S. Krinsky, M. Vijayaraghavan, L. Schneider. Financial ties between DSM-IV panel members and the pharmaceutical industry. *Psychotherapy and Psychosomatics*. Vol. 75, pg. 154-160, 2006. <https://doi.org/10.1159/000091772>.
- 9 L. Cosgrove, S. Krinsky. A comparison of DSM-IV and DSM-5 panel members' financial associations with industry: a pernicious problem persists. *PLoS Medicine*. Vol. 9, pg. e1001190, 2012. <https://doi.org/10.1371/journal.pmed.1001190>.
- 10 A. Lundh, J. Lexchin, B. Mintzes, J. B. Schroll, L. Bero. Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews*. Vol. 2, pg. MR000033, 2017. <https://doi.org/10.1002/14651858.MR000033.pub3>.
- 11 M. R. Liebowitz, J. M. Gorman, A. J. Fyer, D. F. Klein. Social phobia: review of a neglected anxiety disorder. *Archives of General Psychiatry*. Vol. 42, pg. 729-736, 1985. <https://doi.org/10.1001/archpsyc.1985.01790310091013>.
- 12 C. Lane. *Shyness: how normal behavior became a sickness*. Yale University Press, 2008.
- 13 U.S. Food and Drug Administration. Paxil (paroxetine hydrochloride) tablets and oral suspension: final approved labeling, social anxiety disorder indication. NDA 20-031/S-023, Center for Drug Evaluation and Research, 1999.
- 14 D. Healy. Yes: marketing hinders discovery of long-term solutions. *Western Journal of Medicine*. Vol. 175, pg. 364, 2001. <https://doi.org/10.1136/ewj.175.6.364>.
- 15 D. V. Sheehan. No: efforts to relieve human suffering deserve rewards. *Western Journal of Medicine*. Vol. 175, pg. 365, 2001. <https://doi.org/10.1136/ewj.175.6.365>.
- 16 J. Drescher. Out of DSM: depathologizing homosexuality. *Behavioral Sciences*. Vol. 5, pg. 565-575, 2015. <https://doi.org/10.3390/bs5040565>.
- 17 W. T. Carpenter, J. van Os. Should attenuated psychosis syndrome be a DSM-5 diagnosis? *American Journal of Psychiatry*. Vol. 168, pg. 460-463, 2011. <https://doi.org/10.1176/appi.ajp.2011.10121816>.
- 18 K. M. Kulage, A. M. Smaldone, E. G. Cohn. How will DSM-5 affect autism diagnosis? A systematic literature review and meta-analysis. *Journal of Autism and Developmental Disorders*. Vol. 44, pg. 1918-1932, 2014. <https://doi.org/10.1007/s10803-014-2065-2>.
- 19 R. Loomes, L. Hull, W. P. L. Mandy. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*. Vol. 56, pg. 466-474, 2017. <https://doi.org/10.1016/j.jaac.2017.03.013>.
- 20 D. A. Regier, W. E. Narrow, D. E. Clarke, H. C. Kraemer, S. J. Kuramoto, E. A. Kuhl, D. J. Kupfer. DSM-5 field trials in the United States and Canada, part II: test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*. Vol. 170, pg. 59-70, 2013. <https://doi.org/10.1176/appi.ajp.2012.12070999>.
- 21 D. A. Regier, W. E. Narrow, D. E. Clarke, H. C. Kraemer, S. J. Kuramoto, E. A. Kuhl, D. J. Kupfer. DSM-5 field trials in the United States and Canada, part I: study design, sampling strategy, implementation, and analytic approaches. *American Journal of Psychiatry*. Vol. 170, pg. 43-58, 2013. <https://doi.org/10.1176/appi.ajp.2012.12070998>.
- 22 A. V. Horwitz, J. C. Wakefield. *The loss of sadness: how psychiatry transformed normal sorrow into depressive disorder*. Oxford University Press, 2007.
- 23 J. C. Wakefield, M. B. First. Validity of the bereavement exclusion to major depression: does the empirical evidence support the proposal to eliminate the exclusion in DSM-5? *World Psychiatry*. Vol. 11, pg. 3-10, 2012. <https://doi.org/10.1016/j.wpsyc.2012.01.002>.
- 24 P. Zachar, M. B. First, K. S. Kendler. The bereavement exclusion debate in the DSM-5: a history. *Clinical Psychological Science*. Vol. 5, pg. 890-906, 2017. <https://doi.org/10.1177/2167702617711284>.
- 25 B. Levis, A. Benedetti, J. P. A. Ioannidis, Y. Sun, Z. Negeri, C. He, Y. Wu, A. Krishnan, P. M. Bhandari, D. B. Rice, B. D. Thombs. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *Journal of Clinical Epidemiology*. Vol. 122, pg. 115-128, 2020. <https://doi.org/10.1016/j.jclinepi.2020.02.002>.
- 26 I. Hacking. The looping effects of human kinds. In D. Sperber, D. Premack, A. J. Premack (Eds.), *Causal cognition: a multidisciplinary debate*. Clarendon Press, pg. 351-383, 1995.
- 27 I. Hacking. *Rewriting the soul: multiple personality and the sciences of memory*. Princeton University Press, 1995.
- 28 J. Paris. The rise and fall of dissociative identity disorder. *Journal of Nervous and Mental Disease*. Vol. 200, pg. 1076-1079, 2012. <https://doi.org/10.1097/NMD.0b013e318275d285>.
- 29 GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry*. Vol. 9, pg. 137-150, 2022. [https://doi.org/10.1016/S2215-0366\(21\)00395-3](https://doi.org/10.1016/S2215-0366(21)00395-3).
- 30 A. Cipriani, T. A. Furukawa, G. Salanti, A. Chaimani, L. Z. Atkinson, Y. Ogawa, S. Leucht, H. G. Ruhe, E. H. Turner, J. P. T. Higgins, M. Egger, N. Takeshima, Y. Hayasaka, H. Imai, K. Shinohara, A. Tajika, J. P. A. Ioannidis, J. R. Geddes. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*. Vol. 391, pg. 1357-1366, 2018. [https://doi.org/10.1016/S0140-6736\(17\)32802-7](https://doi.org/10.1016/S0140-6736(17)32802-7).
- 31 C. Henderson, G. Thornicroft. Stigma and discrimination in mental illness: Time to Change. *The Lancet*. Vol. 373, pg. 1928-1930, 2009. [https://doi.org/10.1016/S0140-6736\(09\)61046-1](https://doi.org/10.1016/S0140-6736(09)61046-1).
- 32 S. Evans-Lacko, C. Henderson, G. Thornicroft. Public knowledge, attitudes and behaviour regarding people with mental illness in England 2009-2012. *British Journal of Psychiatry*. Vol. 202, pg. s51-s57, 2013. <https://doi.org/10.1192/bjp.bp.112.112979>.
- 33 E. P. Kvaale, N. Haslam, W. H. Gottdiener. The 'side effects' of medicalization: a meta-analytic review of how biogenetic explanations affect stigma. *Clinical Psychology Review*. Vol. 33, pg. 782-794, 2013. <https://doi.org/10.1016/j.cpr.2013.06.002>.
- 34 T. J. Layton, M. L. Barnett, T. R. Hicks, A. B. Jena. Attention deficit-hyperactivity disorder and month of school enrollment. *New England Journal of Medicine*. Vol. 379, pg. 2122-2130, 2018. <https://doi.org/10.1056/NEJMoa1806828>.
- 35 A. M. Ruscio. The latent structure of social anxiety disorder: consequences of shifting to a dimensional diagnosis. *Journal of Abnormal Psychology*. Vol. 119, pg. 662-671, 2010. <https://doi.org/10.1037/a0019341>.

- 
- 36 D. M. Kollman, T. A. Brown, G. I. Liverant, S. G. Hofmann. A taxometric investigation of the latent structure of social anxiety disorder in outpatients with anxiety and mood disorders. *Depression and Anxiety*. Vol. 23, pg. 190-199, 2006. <https://doi.org/10.1002/da.20158>.
- 37 J. R. T. Davidson, D. C. Hughes, L. K. George, D. G. Blazer. The boundary of social phobia: exploring the threshold. *Archives of General Psychiatry*. Vol. 51, pg. 975-983, 1994. <https://doi.org/10.1001/archpsyc.1994.03950120047008>.
- 38 T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow, P. Wang. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*. Vol. 167, pg. 748-751, 2010. <https://doi.org/10.1176/appi.ajp.2010.09091379>.
- 39 R. Kotov, R. F. Krueger, D. Watson, T. M. Achenbach, R. R. Althoff, R. M. Bagby, T. A. Brown, W. T. Carpenter, A. Caspi, L. A. Clark, N. R. Eaton, M. Forbes, K. T. Forbush, D. Goldberg, D. Hasin, S. E. Hyman, M. R. Ivanova, D. R. Lynam, K. Markon, J. D. Miller, T. J. Moffitt, L. C. Morey, S. N. Mullins-Sweatt, J. Ormel, C. J. Patrick, D. B. Samuel, M. Sellbom, L. J. Simms, A. Skodol, T. Slade, S. C. South, J. Tackett, I. Waldman, M. Waszczuk, T. A. Widiger, A. G. Wright, M. Zimmerman. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*. Vol. 126, pg. 454-477, 2017. <https://doi.org/10.1037/abn0000258>.
- 40 D. Borsboom, A. O. J. Cramer. Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*. Vol. 9, pg. 91-121, 2013. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>.
- 41 A. Caspi, R. M. Houts, D. W. Belsky, S. J. Goldman-Mellor, H. Harrington, S. Israel, M. H. Meier, S. Ramrakha, I. Shalev, R. Poulton, T. E. Moffitt. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*. Vol. 2, pg. 119-137, 2014. <https://doi.org/10.1177/2167702613497473>.
- 42 M. B. First, W. Gaebel, M. Maj, D. J. Stein, C. S. Kogan, J. B. Saunders, V. B. Poznyak, O. Gureje, R. F. Lewis-Fernandez, A. Maercker, C. M. Brewin, M. Cloitre, A. Claudino, K. A. Pike, P. Baird, G. M. Reed. An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry*. Vol. 20, pg. 34-51, 2021. <https://doi.org/10.1002/wps.20825>.
- 43 L. Batstra, A. Frances. Diagnostic inflation: causes and a suggested cure. *Journal of Nervous and Mental Disease*. Vol. 200, pg. 474-479, 2012. <https://doi.org/10.1097/NMD.0b013e318257c4a2>.