

PERSPECTIVE

Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it?

S Kapur¹, AG Phillips² and TR Insel³

Patients with mental disorders show many biological abnormalities which distinguish them from normal volunteers; however, few of these have led to tests with clinical utility. Several reasons contribute to this delay: lack of a biological 'gold standard' definition of psychiatric illnesses; a profusion of statistically significant, but minimally differentiating, biological findings; 'approximate replications' of these findings in a way that neither confirms nor refutes them; and a focus on comparing prototypical patients to healthy controls which generates differentiations with limited clinical applicability. Overcoming these hurdles will require a new approach. Rather than seek biomedical tests that can 'diagnose' DSM-defined disorders, the field should focus on identifying biologically homogenous subtypes that cut across phenotypic diagnosis—thereby sidestepping the issue of a gold standard. To ensure clinical relevance and applicability, the field needs to focus on clinically meaningful differences between relevant clinical populations, rather than hypothesis-rejection versus normal controls. Validating these new biomarker-defined subtypes will require longitudinal studies with standardized measures which can be shared and compared across studies—thereby overcoming the problem of significance chasing and approximate replications. Such biological tests, and the subtypes they define, will provide a natural basis for a 'stratified psychiatry' that will improve clinical outcomes across conventional diagnostic boundaries.

Molecular Psychiatry advance online publication, 7 August 2012; doi:10.1038/mp.2012.105

Keywords: clinical tests; diagnosis; stratified medicine; stratified psychiatry

INTRODUCTION

Biological psychiatry aims to understand mental disorders in terms of the biological function of the nervous system. By several measures it has been a tremendous success—thousands of scientific papers and hundreds of books devoted to this subject; legions of dedicated scientists and over 60 dedicated professional societies worldwide; and a profound impact on the public's perception of mental disorders. Despite these successes, it has not led to clinical tests that can be routinely used in the diagnosis and treatment of mental disorders. In the early 2000s, a series of white papers expressed hope that the advances in genetics, imaging and new technologies might lead to a biologically supported psychiatric classification and diagnostic system.¹ But a decade later, as we stand at the threshold of a new version of the DSM, there are few biological clinical tests central to diagnosing psychiatric illnesses (other than those used to exclude physical illnesses). This article explores why this journey has been difficult for psychiatry and what can be done about it.

CLINICAL TESTS IN OTHER MEDICAL SPECIALITIES

All branches of medicine, like psychiatry, began by classifying diseases on the basis of reported symptoms and externally observed clinical signs. However, in the latter half of the nineteenth century, the rest of medicine took a different turn: with the development of the germ-theory and its use for objective tests to demonstrate cause of disease, clinical tests increasingly became central to the practice of medicine.² Thus, patients who in

early classifications would be noted to have 'dropsy and dyspnea' were successively subjected to listening of their murmurs with a stethoscope, to picturing of their enlarged heart in a chest X-ray, to recording of their arrhythmias with an electrocardiogram, to calculation of their ejection fractions with a 2-D Echo and increasingly to a series of new biomarkers (for example, atrial natriuretic factor) that lead to a more refined diagnoses and targeted treatment. As these measures evolved from experimental findings to clinical tests, their ability to predict was demonstrated in real-world settings (clinical validity), and it was demonstrated that those patients who undergo the test, on average, fare better than patients who do not (clinical utility).³ Accordingly, there are over 3000 standardised laboratory diagnostic tests on offer and hundreds of applications of objective diagnostic devices (for example, electroencephalography (EEG), electrocardiogram (EKG) Imaging) in clinical medicine. Few, if any, such tests used in the routine practice of clinical psychiatry.

Although a number of biological findings have been proposed as possible tests in clinical psychiatry—nothing quite caught the attention of the field as the 'Pink Spot' in Schizophrenia in the 1960s,⁴ followed by the Dexamethasone Suppression Test in the 1970s and 1980s. The latter showed initial promise to diagnose endogenous depression with accuracy and predict drug response and clinical relapse.^{5,6} Yet, after thousands of patients were tested an American Psychiatric Association task force concluded⁷ that the test had a rather low sensitivity (40–50% for depression, 60–70% for endogenous forms), modest specificity (often < 70%) and limited clinical utility. This general story of an exciting initial biological finding and claims of a potential test that

¹King's College London, Institute of Psychiatry, London, UK; ²University of British Columbia, Department of Psychiatry and The CIHR Institute of Neurosciences, Mental Health and Addiction, Vancouver, BC, Canada and ³National Institute of Mental Health, Bethesda, MD, USA. Correspondence: Professor S Kapur, King's College London, Institute of Psychiatry, DeCrespigny Park, London SE5 8AF, UK.

E-mail: shitij.kapur@kcl.ac.uk

Received 2 December 2011; revised 21 May 2012; accepted 29 May 2012

subsequently wanes due to limited accuracy or generalizability in real-life clinical settings has been repeated many times.^{8,9} The lack of clinical tests is striking given that biological psychiatry has been very productive in generating new scientific findings: a corpus of over 107 000 articles already available on PubMed with over a 100 new articles being added to this every single week. Why then has it been so difficult to convert biological findings into clinical tests for use in psychiatry?

THE MISSING GOLD STANDARD

To create a biological test to assist in the diagnosis of an illness one needs a stable and biologically valid concept of the illness. Although the International Classification systems for physical illnesses began in 1900, the first efforts to formulate a comprehensive diagnostic scheme in psychiatry did not occur until the 1950s when the International Classification of Diseases-6 (ICD-6) first addressed a chapter dedicated to mental disorders and the first version of the Diagnostic and Statistical Manual of mental disorders, the DSM-I arrived. It was not until the 1970s that serious efforts were made to operationalise the early vignettes into standardised diagnostic criteria and although Guze and Robins¹⁰ anticipated the critical role that laboratory tests might have, they lamented the absence of any such viable tests at that time. In the 1990s, two decades later, when the expected laboratory tests had not arrived, Andreasen¹¹ called for 'new models and new approaches' to diagnostic validation based on genetics and imaging. In the early 2000s, as the stage was being set for the Research Agenda for DSM-V,¹ a similar hope was expressed again—although the promise is yet to be realised a decade later.

As the standardised classification systems have been constantly revised (from ICD-6 to ICD-10 and from DSM-I to DSM-IV), they have remained a descriptive taxonomy based on expressed feelings and observed behaviour. On the one hand, these successive editions of DSM and ICD lead to increasing psychometric precision. On the other hand, the ever increasing fractionation of mental distress into smaller and more numerous categories, without *a priori* biological validity, makes it harder to find specific biomedical tests that diagnose or predict the disorders. The search for specific clinical tests is further complicated by extensive comorbidity across these disorders is rather common. Psychiatric disorders tend to breed across categories almost as frequently as within¹² and their genetic predispositions defy the conventional diagnostic boundaries. Furthermore, the very concept of 'categorical' psychiatric disorders is questioned by some who suggest that a dimensional spectrum may provide a better account of the clinical reality.^{13,14} Even if one acknowledges the primacy of biological factors in some psychiatric disorders, it does not inevitably follow that a biological test would be necessarily most informative or effective in identifying them. Kendler¹⁵ argue that genes and molecules have to work via dozens of 'mechanistic levels' (for example, molecules are embedded in membranes, which form neurons, which form ensembles, which fire in a certain order and so on), therefore the likelihood that a biological alteration will have a powerful one-to-one mapping with a DSM-defined mental disorder may be unlikely.

Thus, psychiatry seems to be in a Catch-22: the current diagnostic system was not designed to facilitate biological differentiation and it does not. The biological studies to date have not been able to propose a clinically viable alternative system. This lack of a gold standard, and the consequent circularity, is not unique to psychiatry. A number of disorders in physical medicine defy simple biological definitions. Breast lumps were categorised based on different symptoms and clinical courses, until histopathological differentiation and molecular markers turned them into distinct illnesses. Arthritides were classified by

symptoms, signs and illness course until immune markers and imaging findings differentiated them into different biologically valid illnesses. If one does not have *a priori* gold standards one can still make progress provided one has biological findings with large effect sizes that correlate with outcomes of a psychiatric disorder. And this has been a challenge for biological psychiatry.

SIGNIFICANCE CHASING WITH UNDERPOWERED STUDIES

The vast majority of biological findings in psychiatry are of a small or moderate effect size—even though many of them survive the ' $P < 0.05$ ' test of statistical significance. Ioannidis has demonstrated that most initial reports of statistically significant but small-effect findings are never substantiated¹⁶ and the ones that are often have even lower effects than initially apparent.¹⁷ Given that efforts to replicate an initial finding usually involve a different clinical setting, a different patient selection and slightly different methods—the chance of replication after an original finding with a $P < 0.05$ is often $< 50\%$.^{18,19} Although these risks are not unique to biological psychiatry, it is particularly vulnerable to 'significance chasing' because the studies in this field generally tend to be underpowered, have small sample sizes,^{20,21} measure multiple dimensions and use subjective outcomes.²²

This challenge of identifying reliable findings on which to base a clinical test strategy is highlighted by two examples: from a handful of articles in the 1970s, there are now over 12 000 articles on 'schizophrenia genetics', with much of this expansion coming in the last decade. The dizzying array of genetic associations is compiled in the SZGene database with over 1700 studies of 1000 genes and 8000 polymorphisms leading to hundreds of 'statistically significant' associations. Collins *et al.*²³ systematically compared some 732 genes implicated in 1374 of these studies, and found that most of these 'findings' were contained in only one study and were never followed up systematically, and the vast majority of these initially positive findings have failed to replicate in subsequent large-scale genome wide analyses.²³ A similar pattern emerges in schizophrenia and imaging—Davidson and Heinrich²⁴ evaluated over two decades of imaging studies in schizophrenia and identified 25 distinct measures amenable to a meta-analysis and found that the majority of these were inconsistent, with more prominent findings associated with greater inter-study inconsistency.^{24,25} This variation is by no means unique to schizophrenia, or genetics or imaging, or even to psychiatry. However, chasing small effects with underpowered studies has meant that even though the field has led to a large output of publications, there are few findings with effect-sizes large enough that could be converted into clinical tests.

APPROXIMATE REPLICATIONS

One might expect that failure to replicate the findings would induce scientists to lose interest in the given area and to move on to findings with more robust effects.²⁶ Unfortunately, an initial underpowered study is often followed by another study of similar size but with a few additional measures and variables to give it some novelty and distinction. These subsequent studies usually have only modest statistical power to decisively confirm or refute the original finding, but do have sufficient multiplicity of new measures to generate some significant finding—even though not precisely the one observed in the first study—thus providing an 'approximate replication'.²⁶ As a result, the 'literature' in the field grows without decisively replicating/rejecting the precise original finding, but instead creates a penumbra of ' $P < 0.05$ ' findings around the first. This problem is well illustrated by the many studies examining frontal dysfunction in schizophrenia. Since the first reports (1998) that 'working memory deficits' are associated with 'frontal dysfunction' in schizophrenia, over 30 studies including 750 individuals have examined this question.²⁷

These studies have used two different imaging technologies, four distinct working memory paradigms and three different modalities (visual, verbal, mixed), with some studies providing a reward, others not, with an average size of a mere 12 subjects. Not surprisingly then, a dozen of these findings show that patients are hyperfrontal as compared with healthy controls, nearly as many show that patients are hypofrontal, whereas a few studies show no discernable difference. There may well be interesting scientific reasons for these opposing findings, perhaps a mediating variable that is yet to be identified, but, until such variation is explained, controlled and removed from such findings, these approximate replications do not provide a reliable basis for clinical tests.

EXTREME COMPARISONS

Nonetheless, some biological findings have stood the test of time and replication, and have reasonably large effect sizes: patients with schizophrenia have larger ventricles and smaller gray matter, their electrophysiological evoked responses are reliably diminished and both pre-pulse inhibition and latent inhibition are impaired.^{21,25} However, these large differences have been noted mostly in studies comparing prototypical patients versus picture-perfect healthy controls. Clinically, one is rarely taxed with distinguishing a textbook patient from a perfectly healthy individual. The real challenge is in distinguishing those who demonstrate the superficially similar symptoms that may merit rather different treatments and outcomes: distinguishing between someone with bipolar depression from unipolar depression, distinguishing someone with severe obsessions with a firm conviction in them from someone with delusions focussed around repetitive behaviors. Experience in the rest of medicine shows that the predictive value of a biological differentiator decreases as we move from extreme contrasts to more clinically relevant ones.^{28,29} Thus it remains unclear whether some of the currently prominent findings would form clinically useful tests, if actually applied to the challenging clinical circumstances in which these tests would actually be useful.

WHAT TO DO ABOUT THIS?

Like the rest of medicine, psychiatry often uses biological tests to exclude other disorders (for example, hypothyroidism in depression, brain tumours in psychosis and so on). But, there are few tests that are used to confirm a diagnosis or a choice of treatment. For successful biological tests one needs important basic biological discoveries. That is a given. But, it is not enough. In addition to basic advances the field needs to be clear about: the kind of tests it seeks, relationship of these putative tests to current diagnostic schemes, standardised ways of collecting and sharing data, and a search for clinically meaningful differences, rather than just statistically significant ones.²⁹

WHAT KIND OF 'TEST' SHOULD WE LOOK FOR?

Screening tests are used to identify diseases in populations that are currently asymptomatic: phenylketonuria in newborns, Pap smears in healthy women. As screening tests are offered to otherwise healthy people—they require very high specificity, must lead to substantial clinical gains and require stringent evaluation of ethical and social implications.^{30,31} Few biological screening tests have been developed without a plausible and understandable link to the aetiology or pathophysiology of the disease—thus biological screening for most psychiatric disorders seems distant. What seems within reach are behavioral screens for early identification and screening for discrete genetic alterations (for example, polymorphisms, copy number variants) associated with a higher risk for behavioral disorders—the opportunities and

complexities of such genetic tests for screening are debated elsewhere.^{32,33}

The most commonly used tests in medicine are those that confirm diagnoses and help choose treatments. The prospects of 'diagnostic tests' for DSM entities remain distant for reasons articulated above, and it seems unlikely that we will replace the 300-disorder taxonomy of the DSM-5 with an alternative biologically based classification system anytime soon. Therefore the real opportunity for psychiatry is to use the emerging advances in genetics, molecular biology, imaging and cognitive science to supplement, rather than replace, the symptom-driven diagnosis. It is often like this in the rest of medicine.

There is currently no single physiological, immunological, histological test for diagnosing asthma—the diagnosis is made based on the pattern of symptoms and clinical findings. Yet, the measurement of forced expiratory volume provides an objective test to determine therapy and monitor the response, and various immunological tests help identify specific aetiologies.³⁴ Arthritis itself remains a clinical diagnosis, but the presence or absence of rheumatoid factor (neither of which is diagnostic or exclusionary of the primary diagnosis) leads to different forms of intervention.³⁵ Thus, while conventional screening and diagnostic tests seem distant—more selective tests that 'subtype' currently prevalent mental disorders or predict potentially beneficial or adverse response to specific drug therapies are within reach.

FROM SUBTYPES TO 'STRATIFIED MEDICINE'—THE PLAUSIBLE GOAL FOR PSYCHIATRY

Ever since Langreth and Waldhoz coined the term 'personalised medicine', authors in psychiatry have enthusiastically endorsed this call—although 'personalized' means different things to different authors.^{36–38} In fact, there are few examples of truly 'personalised' medicine, if by it one means unique intervention customised just for the given individual (quite like a bespoke tailored jacket, such that no two fits are alike). Some emerging patient-personalised vaccines³⁹ or the use of an individual's own cells to derive grafts⁴⁰ are exemplars of truly personalised medicine. It is hard to envisage large-scale application of this principle to psychiatry or medicine any time soon. A more feasible opportunity for psychiatry, as for the rest of medicine, is 'stratified medicine':⁴¹ the identification of biomarkers or cognitive tests that stratify a broad-illness phenotype into a finite number of treatment-relevant subgroups (keeping with the sartorial analogy above, a jacket with a series of chest sizes rather than a one-size-fits-all approach used currently).

Progress in oncology illustrates this approach well: overexpression of human epidermal growth factor subtype 2 (HER2) in breast cancer tissue was first identified as a subtype with a poor prognosis.⁴² As the differential biology of this subtype was better understood, it led to the development of monoclonal antibody therapies (trastuzumab or Herceptin) which increased long-term survival for this particular subtype of breast cancer.⁴³ While HER2 was first observed in breast cancer, overexpression of HER2 has now been observed in subtypes of ovarian, endometrial, non-small-cell lung and gastric cancer and the HER2 stratification is being used to guide treatment in these cancers as well. Several variants of this 'stratified' approach are now making their way to the clinic: the use of K-ras mutations to stratify colorectal cancer, thereby identifying patients who would not benefit from cetuximab; the use of UGT1A1 polymorphisms to identify subgroups of patients who should avoid irinotecan in its treatment.⁴⁴

This approach to 'stratified medicine' has several important lessons for psychiatry: First, it bypasses the nosological debates about the precise diagnostic boundaries and does not need an external 'gold standard', as the approach justifies itself by its utility.⁴⁵ Second, stratification does not require a complete

understanding of aetiology—it was possible to stratify patients based on HER2, even though the ultimate aetiology of breast cancer remains unknown. Third, one does not have to wait for new treatments to arrive—stratification to predict prognosis became possible almost a decade before a viable treatment became available.⁴² Finally, these tests become useful in clinical medicine across diagnoses without requiring wholesale diagnostic reclassification: HER-2 sub-typing is clinically useful in breast, ovarian, lung and gastric cancer, yet, each of these cancers remain distinct clinical entities. Thus, in a ‘stratified psychiatry’, these tests could coexist alongside the conventional diagnostic systems (such as DSM5 or ICD11). The patients could be first diagnosed along conventional grounds, but then stratified by markers that predict prognosis or suggest differential treatments.

The earliest instances of this in psychiatry are already emerging from pharmacogenomics: of the 119 FDA-approved pharmacogenomic biomarkers, which appear in drug labels, 30 of them relate to psychiatric drugs.⁴⁶ Almost all of the psychiatric biomarkers are variants of CYP2D6 and CYP2C19 drug-metabolising enzymes, predicting pharmacokinetic interactions. None of them are indicated for stratifying patients for drug choice or prognosis. Thus, although the principle of biomarkers has officially entered our drug labels, it is yet to make a major therapeutic impact, and for that to happen, biological psychiatry may need to change the way in which studies are done and reported.

STRATIFIED PSYCHIATRY—IMPLICATIONS FOR HOW DATA ARE COLLECTED AND SHARED

For the last three decades, the majority of the grants and papers in biological psychiatry had three characteristics: strict allegiance to a DSM or ICD diagnoses; focus on differentiation of patients with the diagnoses from normal controls; and usually a short-term cross-sectional evaluation. Stratified psychiatry will require a change in this mind-set. The field will have to collect data across the current diagnostic categories, focus on comparing across disorders as much as comparing versus normal controls and will need to collect and curate data, so that it can be widely shared and collated.

The National Institutes of Mental Health, the major funder of biological psychiatry research in the United States of America, has already initiated such an approach when it comes to thinking beyond classical diagnosis. The Research Domain Criteria⁴⁷ is an approach that attempts to link behavior and cognitive symptoms to the underlying neurobiological systems and genetic predispositions in a way that cuts across the categories within the current diagnostic systems. Although the ultimate hope of the Research Domain Criteria is to provide neurobiologically based diagnostic systems with greater validity and reliability, in the nearer term, it may provide a natural basis for ‘subtyping’ one or more disorders in conjunction with existing DSM diagnoses.

Regardless of the diagnostic methodology, most current studies in biological psychiatry include small cross-sectional samples. Snellenberg *et al.*²¹ evaluated 30 clinical studies using a working-memory task and fMRI in schizophrenia, and observed a median sample size of 12;²⁷ Allen *et al.*²¹ examined 13 studies of lateral ventricular volume in patients with schizophrenia—median patient sample 21; similarly, 15 studies of pre-pulse inhibition in schizophrenia had a median clinical sample size of 25. In theory, these several small studies could be combined into conclusive meta-analyses. In practice, almost every meta-analysis of such findings concludes that the myriad of technical differences in the protocols applied, differences in patients selected and the diverse outcomes measured, make it impossible to derive a precise quantitative conclusion.^{21,25} Although our current methods are sufficient to highlight the areas of potential

interest—they do not provide the precision required for clinical test development.

Fortunately, this situation is changing. After a decade of underpowered studies linking single-nucleotide polymorphisms in candidate genes to given psychiatric disorders, the recent Psychiatric GWAS Consortium brought together 160 investigators, from 65 institutions in 19 countries, to pool data across five major psychiatric disorders, encompassing nearly 50 000 patients for what is without doubt the single largest effort in biological psychiatry.⁴⁸ The Consortium required not only aligning the incentives of individual scientists, but also the harmonisation of the already-collected genetic and phenotypic data, and the deposition of data in controlled-access repositories to ensure continuing future use by the original participants and beyond. Similarly, after hundreds of studies using fMRI with different methods and sequences, the 1000 Connectomes Project, an international partnership between 35 laboratories in 10 countries, have shown the tremendous power of standardizing, integrating and sharing data from over 1400 subjects to establish the human ‘functional connectome’.⁴⁹ The Connectomes project highlighted the major variability across centres, but also provided a means for partialling out these effects. Likewise, after nearly half a century of use of diverse instruments to measure cognition in schizophrenia, all of which have documented some form of cognitive dysfunction in schizophrenia, a consensus battery has been developed with input of scientists, the industry and the regulators to allow for a standardised assessment of cognition, not just in the English-speaking world, but in many of the other major languages.⁵⁰

STRATIFIED PSYCHIATRY—IMPLICATIONS FOR HOW STUDIES ARE REPORTED

More standardised studies will not by themselves lead to clinical tests unless the field makes meaningful clinical difference, rather than minimal hypothesis rejection, its priority. A search for ‘meaningful differences’ would require a shift from *P*-values to effect sizes in our scientific discourse.^{29,51} Effect sizes convey the magnitude of a difference, which is easy to comprehend and relates more directly to clinical relevance;⁵² an effect size of 0.2 is only likely to lead to tests which differentiate 10% of the patient population (for example, 55% of the patients will be positive, so will 45% of controls); a medium effect size (0.5) is likely to deliver a 24% differentiation (for example, 62% patients will be positive versus 38% of controls) and only large effect sizes, say 2, provide differentiations of 70% (85% of patients will be positive versus only 15% of controls).^{53,54} The actual differentiation depends not only on effect sizes but also on the expected number of patients versus controls in the relevant clinical setting. Thus the field should demand of its authors not just *P*-values and effect sizes, but, estimates of positive and negative predictive values assuming realistic clinical contexts (see Perlis²⁹ for a recent thoughtful review of options). Such a requirement will keep authors from making superficial claims about the possibility of a ‘clinical test’ when the possibility is minimal,²¹ and this would allow the field to focus on the few possibilities that are likely to yield useful clinical tests instead of the many that surely will not.

STRATIFIED PSYCHIATRY—NOT JUST ABOUT BIOLOGICAL TESTS

Although we have drawn on examples that have used genes and molecules as stratifiers, the most effective ‘stratifiers’ in psychiatry may well come from standardized cognitive and psychological measures. Whether such stratifiers are considered ‘biological’ or not is a semantic debate—what will be critical is that they enhance the ability to understand, predict and prognosticate beyond conventional DSM diagnoses. Psychiatry is likely to be in a position

where it might have to rely on a combination of such tests—some biological, some cognitive, some psychological—to reach effective stratification, and will have to develop sophisticated techniques to identify the ‘additional predictive value’ of such supplementary tests. These issues are not unique to psychiatry—although our starting point may be different. The National Academies of Sciences has recently issued a report calling for a revision of the taxonomy of all diseases based on the emerging new molecular information and going beyond the traditional emphasis on ‘signs and symptoms’.⁵⁵ Moreover, in areas like breast cancer—where there is now a surfeit of predictive stratifiers of different types: tumour size, lymph node involvement, histopathological type, estrogen and HER-2 expression on the tumor and a 70-marker MammaPrint arrays—the field is grappling with how to optimally combine these different stratifiers in way that provides optimal clinical utility.⁵⁶

CONCLUSION

Biological psychiatry and the related neurosciences have changed mankind’s view of itself and of mental illness, but have yet to provide biomedical tests for routine clinical practice. The delay is understandable given the later start than the rest of medicine, the complexity of the brain, the nascence of neuroscientific techniques and the evolving nature of psychiatric nosology. On the other hand, the opportunity afforded by the progress in genomics and imaging combined with the computational abilities is unprecedented and could deliver useful clinical tests. These tests will identify homogenous populations for whom one could develop targeted new therapeutics thus realising a vision of a new stratified psychiatry that cuts across the traditional diagnostic boundaries while simultaneously transforming them.

CONFLICT OF INTEREST

SK has received grant support from GSK and has served as consultant and/or speaker for AstraZeneca, Bioline, BMS-Otsuka, Eli Lilly, Janssen (J&J), Lundbeck, NeuroSearch, Pfizer, Roche, Servier and Solvay Wyeth. AGP serves on the Board of Allon Therapeutics Inc., and holds shares in this corporation. TI has no financial interests to disclose.

ACKNOWLEDGEMENTS

We would like to thank Dr Bruce Cuthbert for his useful comments on an earlier version of this manuscript. SK’s research related to the article is supported by G0701748/1 from the MRC and the Innovative Medicines Initiative (IMI) grant NEWMEDS, under Grant Agreement N8 115008. SK received salary support from the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London.

REFERENCES

- 1 Kupfer D, First M, Regier D (eds). *A Research Agenda for DSM-V*. American Psychiatric Association: Washington, DC, 2002.
- 2 Anderton D. Disease, concepts and classification of. In: Demeny P, McNicol G (eds). *The Encyclopedia of Population*, vol. 1 Macmillan Reference: New York, 2003, pp 247–250.
- 3 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003; **49**: 1–6.
- 4 Editorial. Lessons of the ‘pink spot’. *Br Med J* 1967; **1**: 382–383.
- 5 Carroll BJ, Curtis GC, Mendels J. Neuroendocrine regulation in depression. II. Discrimination of depressed from nondepressed patients. *Arch Gen Psychiatry* 1976; **33**: 1051–1058.
- 6 Goldberg IK. Dexamethasone suppression tests in depression and response to treatment. *Lancet* 1980; **2**: 92.

- 7 The dexamethasone suppression test: an overview of its current status in psychiatry. The APA Task Force on Laboratory Tests in Psychiatry. *Am J Psychiatry* 1987; **144**: 1253–1262.
- 8 Loosen PT, Garbutt JC, Prange AJ. Evaluation of the diagnostic utility of the TRH-induced TSH response in psychiatric disorders. *Pharmacopsychiatry* 1987; **20**: 90–95.
- 9 Nuwer MR. On the controversies about clinical use of EEG brain mapping. *Brain Topogr* 1990; **3**: 103–111.
- 10 Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry* 1970; **126**: 983–987.
- 11 Andreasen NC. The validation of psychiatric diagnosis: new models and approaches. *Am J Psychiatry* 1995; **152**: 161–162.
- 12 Dean K, Stevens H, Mortensen PB, Murray RM, Walsh E, Pedersen CB. Full spectrum of psychiatric outcomes among offspring with parental history of mental disorder. *Arch Gen Psychiatry* 2010; **67**: 822–829.
- 13 Allardyce J, Suppes T, Van Os J. Dimensions and the psychosis phenotype. *Int J Methods Psychiatr Res* 2007; **16**(Suppl 1): S34–S40.
- 14 Andrews G, Brughla T, Thase ME, Duffy FF, Rucci P, Slade T. Dimensionality and the category of major depressive episode. *Int J Methods Psychiatr Res* 2007; **16**(Suppl 1): S41–S51.
- 15 Kendler KS. Explanatory models for psychiatric illness. *Am J Psychiatry* 2008; **165**: 695–702.
- 16 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; **2**: e124.
- 17 Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008; **19**: 640–648.
- 18 Cumming G. Replication and *p* intervals: *p* values predict the future only vaguely, but confidence intervals do much better. *Perspect Psychol Sci* 2008; **3**: 286–300.
- 19 Miller J. What is the probability of replicating a statistically significant effect? *Psychon Bull Rev* 2009; **16**: 617–640.
- 20 Rothpearl AB, Mohs RC, Davis KL. Statistical power in biological psychiatry. *Psychiatry Res* 1981; **5**: 257–266.
- 21 Allen AJ, Griss ME, Folley BS, Hawkins KA, Pearlson GD. Endophenotypes in schizophrenia: a selective review. *Schizophr Res* 2009; **109**: 24–37.
- 22 Uher R. Gene-environment interaction: overcoming methodological challenges. *Novartis Found Symp* 2008; **293**: 13–26; discussion 26–30, 68–70.
- 23 Collins A, Kim Y, Sklar P, O’Donovan M, Sullivan P. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol Med* 2012; **42**: 607–616.
- 24 Davidson LL, Heinrichs RW. Quantification of frontal and temporal lobe brain-imaging findings in schizophrenia: a meta-analysis. *Psychiatry Res* 2003; **122**: 69–87.
- 25 Heinrichs RW. Meta-analysis and the science of schizophrenia: variant evidence or evidence of variants? *Neurosci Biobehav Rev* 2004; **28**: 379–394.
- 26 Maxwell SE. The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol Methods* 2004; **9**: 147–163.
- 27 Van Snellenberg JX, Torres IJ, Thornton AE. Functional neuroimaging of working memory in schizophrenia: task performance as a moderating variable. *Neuropsychology* 2006; **20**: 497–510.
- 28 Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978; **299**: 926–930.
- 29 Perlis RH. Translating biomarkers to clinical practice. *Mol Psychiatry* 2011; **16**: 1076–1087.
- 30 Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet* 2002; **359**: 881–884.
- 31 Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N *et al*. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP Working Group. *Gen Med* 2009; **11**: 3–14.
- 32 Grosse SD, Rogowski WH, Ross LF, Cornel MC, Dondorp WJ, Khoury MJ. Population screening for genetic disorders in the 21st century: evidence, economics, and ethics. *Public Health Genomics* 2010; **13**: 106–115.
- 33 Hoge SK, Appelbaum PS. Ethics and neuropsychiatric genetics: a review of major issues. *Int J Neuropsychopharmacol* 2012; 1–11; PMID: 22372758.
- 34 Gibson PG, McDonald VM, Marks GB. Asthma in older adults. *Lancet* 2010; **376**: 803–813.
- 35 Scott DL, Wolfe F, Huizinga TWJ. Rheumatoid arthritis. *Lancet* 2010; **376**: 1094–1108.
- 36 Gurwitz D, Weizman A. Personalized psychiatry: a realistic goal. *Pharmacogenomics* 2004; **5**: 213–217.
- 37 Brammer M. The role of neuroimaging in diagnosis and personalized medicine—current position and likely future directions. *Dialogues Clin Neurosci* 2009; **11**: 389–396.
- 38 de Leon J. The future (or lack of future) of personalized prescription in psychiatry. *Pharmacol Res* 2009; **59**: 81–89.
- 39 Belli F, Testori A, Rivoltini L, Maio M, Andreola G, Sertoli MR *et al*. Vaccination of metastatic melanoma patients with autologous tumor-derived heat shock protein gp96-peptide complexes: clinical and immunologic findings. *J Clin Oncol* 2002; **20**: 4169–4180.

- 40 Kocher AA, Schuster MD, Szabolcs MJ, Takuma S, Burkhoff D, Wang J *et al*. Neovascularization of ischemic myocardium by human bone-marrow-derived angioblasts prevents cardiomyocyte apoptosis, reduces remodeling and improves cardiac function. *Nat Med* 2001; **7**: 430–436.
- 41 Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev Drug Discov* 2007; **6**: 287–293.
- 42 Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human-breast cancer—correlation of relapse and survival with amplification of the her-2 neu oncogene. *Science* 1987; **235**: 177–182.
- 43 Smith I, Procter M, Gelber RD, Guillaume S, Feyereislova A, Dowsett M *et al*. 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. *Lancet* 2007; **369**: 29–36.
- 44 Ferraldeschi R, Newman WG. Pharmacogenetics and pharmacogenomics: a clinical reality. *Ann Clin Biochem* 2011; **48**(Part 5): 410–417.
- 45 Hyman SE. Opinion—can neuroscience be integrated into the DSM-V? *Nat Rev Neurosci* 2007; **8**: 725–U716.
- 46 FDA U. Table of Pharmacogenomic Biomarkers in Drug Labels. Available at <http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm> (accessed 1 December 2011).
- 47 Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K *et al*. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* 2010; **167**: 748–751.
- 48 Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. *Neuron* 2010; **68**: 182–186.
- 49 Biswal BB, Mennes M, Zuo XN, Gohel S, Kelly C, Smith SM *et al*. Toward discovery science of human brain function. *Proc Natl Acad Sci USA* 2010; **107**: 4734–4739.
- 50 Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S *et al*. Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICES conference to select cognitive domains and test criteria. *Biol Psychiatry* 2004; **56**: 301–307.
- 51 Kraemer HC. DSM categories and dimensions in clinical and research contexts. *Int J Methods Psychiatr Res* 2007; **16**: S8–S15.
- 52 Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev* 2007; **82**: 591–605.
- 53 Rosenthal R, Rubin DB. A simple, general-purpose display of magnitude of experimental effect. *J Educ Psychol* 1982; **74**: 166–169.
- 54 Coe R. It's the effect size, stupid: what effect size is and why it is important. A paper presented at the Annual Conference of the British Educational Research Association 2002.
- 55 Committee on a Framework for Developing a New Taxonomy of Disease of the National Academies of Sciences U. (ed) *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. The National Academies Press: Washington, DC, 2011.
- 56 Reis-Filho JS, Puzstai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 2011; **378**: 1812–1823.