

The Hamilton Rating Scale for Depression: The making of a “gold standard” and the unmaking of a chronic illness, 1960–1980

Chronic Illness

9(3) 202–219

© The Author(s) 2012

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1742395312467658

chi.sagepub.com**Michael Worboys****Abstract**

Objectives: To show why and how the Hamilton Rating Scale for Depression became the ‘Gold Standard’ for assessing therapies from the mid-1960s and how it was used to frame depression as a short-term and curable illness rather than a chronic one.

Methods: My approach is that of the social construction of knowledge, identifying the interests, institutional contexts and practices that produce knowledge claims and then mapping the social processes of their circulation, validation and acceptance.

Results: The circulation and validation of Hamilton Rating Scale for Depression was relatively slow and it became a ‘Gold Standard’ ‘from below’, from an emerging consensus amongst psychiatrists undertaking clinical trials for depression, which from the 1960s were principally with psychopharmaceuticals for short-term illness. Hamilton Rating Scale for Depression, drug trials and the construction of depression as non-chronic were mutually constituted.

Discussion: Hamilton Rating Scale for Depression framed depression and its sufferers in new ways, leading psychiatrists to understand illness as a treatable episode, rather than a life course condition. As such, Hamilton Rating Scale for Depression served the interests of psychiatrists and psychiatry in its new era of drug therapy outside the mental hospital. However, Hamilton Rating Scale for Depression was a strange kind of ‘standard’, being quite non-standard in the widely varying ways it was used and the meanings given to its findings.

Keywords

Depression, clinical scales, psychopharmaceuticals, chronic illness, standards

Received 26 July 2012; accepted 4 October 2012

Introduction

There has been much discussion in recent years about whether depression is a chronic illness against the modern view that it is

Centre for the History of Science, Technology and Medicine, University of Manchester, Manchester, UK

Corresponding author:

Michael Worboys, Centre for the History of Science, Technology and Medicine, University of Manchester, Simon Building, Manchester M13 9PL, UK.

Email: michael.worboys@manchester.ac.uk

typically time-limited.¹ Gask dated the growing dominance of this view to the 1980s and ‘the launch and promotion of a new group of antidepressants, the selective serotonin reuptake inhibitors (SSRIs)’.² The traditional view of depressive illness, from melancholia in the nineteenth century to Kraepelin’s characterisation of manic depression that dominated twentieth-century psychiatry, was that the illness was recurrent, chronic or both. Sufferers could spend years in mental hospitals, where, from the 1950s, they might receive regular electroconvulsive therapy (ECT). The changes in the last quarter of the twentieth century are well known and recognised as revolutionary at all levels: definitions of ‘depression’ and the impact of DSM-III; the treatment of choice shifting from ECT to drugs; the closure of long-stay hospitals and the development of community care where sufferers from depression are mainly treated by general practitioners. The impact of these changes on medical views of depression was evident in an Editorial in *Psychological Medicine* in 2012, which had to remind readers of new evidence that amongst patients diagnosed with depression, only half had a single episode and half had a recurrent and chronic life-long illness.³ The authors argued that more effort should now be given to identifying recurrence, with a view to altering ‘the trajectory of depression that is so chronic, severe and disabling’ for ‘the betterment of so very many’.

Methods

My principal research question is when and how did the view that depression was typically time-limited and non-chronic originate? Was it in the 1980s and early 1990s with the arrival of SSRIs? These drugs were undoubtedly important, but so too were the changes in service provision and a host of other patient, professional and other

factors. In this article, I investigate the longer-term origins in ways that depression was framed by psychiatrists through the impact of the Hamilton Rating Scale for Depression (HRSD), which from the 1970s became, and to a large extent remains, the dominant tool in assessing the severity of depression. A key feature of HRSD was that it was used to measure the outcome of treatment, especially drugs, and was applied as a ‘before and after’ schema, leading to the view that depression was event, thereby downplaying seriality. My argument also offers a case study of the impact of standard scales in medicine, and the interaction of drug standards and standard drugs.

My methods are those of the social construction of knowledge, explaining how ways of knowing and practising are formulated in specific social contexts, then circulated and validated in contingent settings by a variety of actors. Constructivist historical methods were applied to articles and books that discussed the application of HRSD to various patient groups in hospital and community setting from the 1960s to the late 1970s. Sources were identified from the standard online databases—Pubmed (keyword) and Science Direct (full text)—and quantitative indicators were derived from Web of Science. Detailed qualitative analysis of selected articles was also made, using close reading to identify the assumptions and modes of analysis of the authors.

Results

The 21-item HRSD for assessing the severity of depression was developed by the English psychiatrist Max Hamilton and presented to the psychiatric community in 1960 in the, then somewhat obscure, *Journal of Neurology, Neurosurgery and Psychiatry*.⁴ Interviewed in 1982, Hamilton observed

that, after completing a number of clinical trials on new drugs,

I was also interviewing people about my depression scale and trying to see if I could get some work going on depression. I went around with my scale and it created a tremendous wave of apathy. They all thought I was a bit mad. Eventually I got it published in the *Journal of Neurology, Neurosurgery and Psychiatry*. It was the only one that would take it.⁵

He took some pleasure in adding that, 'And now everyone tells me the scale is wonderful, I always remember when it had a different reception. This makes sure I don't get a swollen head.' Whether the last point was accurate is open to debate, as Hamilton was quite a domineering figure, but there is no doubt that his rating scale was, and still is, widely used. It has earned the title of the 'Gold Standard' for the assessment of depression, though its reign may now be limited.⁶ Given its status and influence, it is surprising that it has not been subject of historical enquiry and even authors who are critical of modern psychiatry and its 'manufacturing of depression' have not subjected it to scrutiny.⁷

There are two explanations of its dominance, both of which have some merit but are not the whole story. The first, which is common amongst psychiatrists, is that HRSD became the 'Gold Standard' simply by being the earliest scale to enjoy widespread use. However, it was born into a world of already competing scales, so the key question to answer is, why and how did it see off its rivals? Interestingly, Hamilton's Anxiety Scale, which was actually published before HRSD and hence was more of a 'first,' did not endure. The second explanation is that HRSD was ideally suited to measure the effects of drug treatments, especially tricyclics such as imipramine, which were 'somewhat anxiolytic and somewhat sedative in effect.'⁸ HRSD scored for

sleep and for weight gain, which were known to be affected by tricyclics. In other words and to quote one reviewer of *The Antidepressant Era*, 'The early drugs defined the very scale that was used to measure their performance.'⁹ One recent critic of the scale wrote that Hamilton 'fashioned his test to meet the needs of his drug company patrons.'⁷ Healy says that there is no evidence that Hamilton used his own scale in clinical practice, but then it was a research rather than clinical tool, designed to quantify changes in a patient's condition over time.¹⁰ It is unclear whether Hamilton had direct 'drug company patrons,' though he was the founding President of the British Association of Psychopharmacology and an early member of the International College of Neuro-Psychopharmacology (CINP), which since his death in 1988 has awarded an annual prize in his name. On the other hand, Hamilton is widely described as an iconoclast and seems to have been a socialist; he was certainly a strong defender of the National Health Service in the 1980s when it was under threat from Thatcher era cuts in public spending. What is clear is that in the late 1950s and early 1960s Hamilton had many motives and that his abrasive character meant that pleasing anyone was not high on the list.

In this article, I argue that the dominance of HRSD was only slowly achieved and that in its first two decades it had many rivals and that no one was more surprised than Hamilton himself that it proved to be so successful. Also, its dominance was largely in clinical research, translating trial findings, quite often, into simple before-and-after scores. There was an inherent bias to consider depression as time-limited and all the more so as a result of drug treatment. Hamilton created the scale to enable psychiatrists to chart changes in already diagnosed patients through particular treatment regimes, converting qualitative judgments into quantitative data on a

fine-grained 100-point scale. The scale also allowed psychiatrists to determine what the most significant changes were in an array of symptoms; though as I will show, most early studies used the aggregated scores rather than disaggregated data. Indeed, studies in the 1980s demonstrated that the schema was modified promiscuously, with psychiatrists adding and subtracting items to assess.¹¹ In 1990, Zitman et al. surveyed five major journals over a year for research papers using the HAM-D and asked authors of for a copy of the scale they used. Fewer than half the investigators referenced the correct version of the HAM-D, and only 4 out of 51 responders used versions that were the same as a published version.

HRSD was not designed as a diagnostic schema, though many used it as such and one reason for its success was that its approach anticipated the emphasis of symptoms and disease entities enshrined in DSM-III in 1980.¹² Although invented well before even DSM-II (1968), Hamilton's scale was for a specific condition and proposed standardisation around overt symptoms, the features that distinguished the third from the second version of the DSM. Shaped by the assumptions of dominant psychodynamic approaches, DSM-I and -II had 'conceived of symptoms as reflections of broad underlying dynamic conditions... that only became meaningful through exploring the personal history of each individual'.¹² Influenced strongly by Karl Menninger's assumption that all mental disorders were reducible to 'the failure of the suffering individual to adapt to his or her environment', psychiatrists tended to focus on finding underlying mental causes and to interpret these as constitutional and likely to be chronic.¹³ DSM-III's move towards specific diseases and to focus on symptoms rather than underlying causes weakened these imperatives.

Max Hamilton was born in Offenbach, near Frankfurt, in 1912, and his parents

moved to London in 1915.¹⁴ He qualified in medicine at University College Hospital London in 1934 and worked in a number of posts before settling upon psychiatry in 1946, when he joined the Maudsley Hospital in London. He worked at various London hospitals and began an association with Cyril Burt that led him to develop expertise in, and an almost missionary commitment to, psychometrics, which was fashionable in the psychological sciences in the 1950s. In 1953, he moved to the University of Leeds as lecturer in psychiatry. He found little time for research and in 1957 resigned to take up a temporary, 2-year research position in the University. This was funded by research grants from the Mental Health Research Trust and by a trial that his head of department, Ronald Hargreaves, was running on chlorpromazine. In this work, Hamilton developed a number of scales, the first in 1957 in a study with Hargreaves on the value of Benactyzine in the treatment of anxiety, for which drugs and placebos were supplied by Glaxo.¹⁵ The anxiety scale, later termed HAMA, anticipated many of the features of HRSD.

We therefore classified all the symptoms likely to be found in our patients under the following headings: (1) anxious mood; (2) tension; (3) specific fears and phobias; (4) sleep disturbance; (5) intellectual disturbance; (6) depressive features; (7) somatic disturbances (muscular and sensory); (8) cardiovascular disturbance; (9) respiratory disturbances; (10) gastro-intestinal disturbances; (11) genitourinary disturbances; (12) autonomic disturbances and (13) manifestations of anxiety in the behaviour at the interview. A gloss was prepared listing the features to be taken into account in making an assessment under any of these headings. At the interview we rated each of these thirteen items on a five-point scale as follows: 0, none; 1, mild; 2, moderate; 3, severe; 4, grossly disabling. This rating

scale yields a variety of different types of information for each patient, including a "profile" of his symptomatology and, by summing the ratings for all headings, a gross symptom score.¹⁵

One conclusion of this study was that 'impressionistic global judgments of a patient's condition alone are of little value in assessing the effect upon him of a particular regime'. Hamilton had previously spoken on the use of scales in this work on anxiety at the British Psychological Society in 1956.¹⁶ In what became a feature of his publications of scales, he devoted much of the paper to sophisticated statistical testing of reliability and reproducibility. As noted already, HAMA was further elaborated in the 1960s but did not have the success of HRSD, but that is a topic for another paper.

The first iteration of the HRSD scale was actually published in 1959, in an article co-authored with Jack White, a consultant psychiatrist at the Stanley Royd Hospital, Wakefield.¹⁷ The famous 1960 paper was already in press and mentioned, though without a citation. The scale in the 1959 paper offered a different and more finely grained classification of patient symptoms, moving away from the three accepted dichotomies: Reactive – Endogenous; Agitated – Retarded; Neurotic – Psychotic. Hamilton and White subjected patient's scores on their schema to factor analysis and identified four groups of patients and types of depression: Endogenous, Doubtful Endogenous, Doubtful Reactive and Reactive. In other words, they were using the scores for the classification of different types of depression. In conclusion, they argued that, with the range of therapeutic options increasing as new drugs were added to ECT and psychotherapy, it was important for psychiatrists to be better able to differentiate forms of depression and their response to treatments. The study was of 64 male patients at Stanley Royd and included an Appendix of case histories of 20 patients,

which showed that they had received a variety of treatments. Of the 20, 16 had received ECT, so the origins of HRSD lie in charting the dominant therapeutic regimes of the era and were not only developed for pharmaceutical treatment.

What became known as HRSD was proposed by Hamilton in his now famous and much cited 1960 paper? His stated aim was to improve upon existing scales, which he criticised for being inappropriate, unreliable or using ill-defined symptoms.⁴ His new scale was to be used in interviews conducted by psychiatrists and was intended for patients already diagnosed with depression. It relied mostly on the observations of bodily (somatic) and behavioural features by psychiatrists, which were also weighted more heavily than the few symptoms that relied on patient's reports of their feelings (Figure 1).

The empirical basis of the paper was drawn from 49 of the 64 patients discussed in the 1959 paper. There were 17 variables in the new scale, each rated on either a four- or two-point range, which produced a potential maximum of 50 points for extremely severe illness. The recommendation was that two psychiatrists interview the patient separately and their scores be added together to give a rating out of 100 (Figure 2). The correlation between the scores of the two scorers (presumably Hamilton and White) was found to be high and to improve with experience.

In discussing individual patients, Hamilton did not use their overall rating score; instead he gave their pattern of factor measures in terms of the four diagnostic groups identified in the 1959 paper with White: Factor 1: Endogenous, Factor 2: Doubtful Endogenous, Factor 3: Doubtful Reactive and Factor 4: Reactive.¹⁷ Figure 3 presents the description of one of the patients whose profile was predominantly Factor 1 and this ends with the classification of his illness as 'endogenous' and seemingly chronic and likely to relapse.

J. Neurol. Neurosurg. Psychiat., 1960, 23, 56.

A RATING SCALE FOR DEPRESSION

BY

MAX HAMILTON

From the Department of Psychiatry, University of Leeds

The appearance of yet another rating scale for measuring symptoms of mental disorder may seem unnecessary, since there are so many already in existence and many of them have been extensively used. Unfortunately, it cannot be said that perfection has been achieved, and indeed, there is considerable room for improvement.

appear in different settings. Other symptoms are difficult to define, except in terms of their settings, e.g., mild agitation and derealization. A more serious difficulty lies in the fallacy of naming. For example, the term "delusions" covers schizophrenic, depressive, hypochondriacal, and paranoid delusions. They are all quite different and should be clearly distinguished. Another difficulty may be

Figure 1. Hamilton's now famous paper on rating scales for depression was published in a little known journal.⁴

APPENDIX I ASSESSMENT OF DEPRESSION

Item No.	Score Range	Symptom	Score
1	0-4	Depressed mood	
2	0-4	Guilt	
3	0-4	Suicide	
4	0-2	Insomnia, initial	
5	0-2	.. middle	
6	0-2	.. delayed	
7	0-4	Work and interests	
8	0-4	Retardation	
9	0-2	Agitation	
10	0-4	Anxiety, psychic	
11	0-4	.. somatic	
12	0-2	Somatic symptoms, gastrointestinal	
13	0-2	.. general	
14	0-2	Genital symptoms	
15	0-4	Hypochondriasis	
16	0-2	Loss of insight	
17	0-2	.. weight	
18	0-2	Diurnal variation { M	
19	0-4	Depersonalization, etc.	
20	0-4	Paranoid symptoms	
21	0-2	Obsessional symptoms	

Grading	
0	Absent
1	Mild or trivial
2	Moderate
3	
4	Severe

0	Absent
1	Slight or doubtful
2	Clearly present

Figure 2. The first published iteration of what became HAM-D or HRSD.⁴

Hamilton made clear the importance of factor scores and their value over the classical clinical categories. In summary, he wrote:

A rating scale is described for use in assessing the symptoms of patients diagnosed as suffering from depressive states. The first four latent vectors of the intercorrelation matrix obtained from 49 male patients are of interest, as shown by (a) the factor saturations, (b) the case histories of patients

scoring highly in the factors and (c) the correlation between factor scores and outcome after treatment. The general problem of the relationship between clinical syndromes and factors extracted from the intercorrelations of symptoms is discussed.⁴

There is no evidence in the paper that 'before and after' treatment scores were taken, the only link to treatment seems to be that the initial factor scores were indicative of the outcome of (mostly ECT)

Factor 1.—A man aged 39 years (Case 39) had factor scores of F_1 76, F_2 37, F_3 49, and F_4 52.

This patient was admitted to hospital after two attempts at suicide, first by electrocution, and, when this failed, by an overdose of phenobarbitone. No psychological precipitating factors were found. On admission he was severely depressed and still actively suicidal. He had strong feelings of guilt, and feared that he had acquired venereal disease and was infecting others with it. He was markedly retarded and showed loss of insight. His sleep was disturbed in all three phases, he had no interest in anything and had complete loss of libido since the onset of his illness four months previously. His symptoms cleared with six courses of electroshock treatment (E.C.T.). Two weeks later he suddenly relapsed and attempted to cut his wrists with a broken tumbler. He again recovered with a further course of E.C.T. and has remained well ever since.

This case was one of classical endogenous depression.

Figure 3. An example of the case histories and commentaries included in Hamilton's 1960 paper.⁴

treatment, hence, this first presentation of HRSD can be read as offering a more refined diagnosis or prognosis. In another paper with Jack White, also published in 1960, Hamilton assessed ratings as an indicator of the outcome of depression treated with ECT.¹⁸

The first published trial to use HRSD was a study of the use of the new drug amitriptyline by CG Burt and colleagues at the Royal Park Hospital in Melbourne, Australia.¹⁹ For each patient an aggregate score out of 50 was first used to group patients; there was no factor analysis.

After initial evaluation on Hamilton's (1960) scale for quantifying depressive illnesses, patients were allocated to one of four groups delineated on the basis of two leading prognostic criteria, age and severity of illness. "Mild young" depressives were aged between 30 and 49 and, out of a possible maximum score of 100, had total scale scores below 40; "young severe" depressives were between 30 and 49 and had total scale scores above 40. Similar criteria of severity were used in the "old

mild" and "old severe", who were aged between 50 and 70.

The same overall rating score was used to assess the outcome after one and then four weeks treatment with amitriptyline compared to imipramine; the latter being the market leader for severe depression. The Table and Chart below show the range in individual rating scores and aggregates for the 'old severe' group. In fact, this was one of the few studies in the period that presented the symptom scores separately, typically the single aggregate score out of 50 or 100 was used (Figure 4).

In their discussion, Burt et al. made two key points about the HRSD that were, and are still, widely stated to account for its widespread use: (1) it was 'simple to use and rapidly completed' and (2) it could map changes that drugs brought in specific symptoms. Burt and his colleagues wrote of 'target' symptoms, which was perhaps an implicit comparison to the blunderbuss of ECT and its impact on the whole psyche. HRSD could certainly also map the

TABLE II
Mean Improvement Scores on Physicians' Scale Ratings of Total Group

Specific Depressive Symptom	Possible Range of Scores	After 1 Week		After 4 Weeks	
		Amitrip. n=37	Imp. n=36	Amitrip. n=37	Imp. n=36
Depressed Mood	0-8	1.4	1.2	3.2	2.4
Retardation	0-8	.7	.7	2.0	1.7
Works & Interests	0-8	1.9	1.8	3.6	2.9
Guilt	0-8	.8	.8	2.1	1.5
Suicide	0-8	2.1	2.1	2.9	2.5
Psychic Anxiety	0-8	1.6	1.6	2.8	2.4
Somatic Anxiety	0-8	1.8	1.1	1.9	2.0
Hypochondriasis	0-8	.7	1.0	1.4	1.1
Agitation	0-4	1.2	1.2	1.7*	1.2
Initial Insomnia	0-4	1.1	.8	2.2*	1.7
Middle Insomnia	0-4	.7*	-.1	1.3	.8
Delayed Insomnia	0-4	.9†	-.3	1.6*	.7
Somatic Symptoms (Gastro-Int.)	0-4	1.2	.9	2.4†	1.5
Somatic Symptoms (General)	0-4	.8	1.0	1.8	1.5
Genital Symptoms	0-4	.9	.7	1.6*	.9
Loss of Weight	0-4	.0	.0	1.9	1.6
Loss of Insight	0-4	.3	.1	.7	.3
Total Improvement Score ..	0-100	17.9	14.7	35.2*	26.7

* Significant at 5 per cent. level.
† Significant at 1 per cent. level.

MEAN ONE-WEEK IMPROVEMENT - "OLD SEVERE" GROUP (n=35)

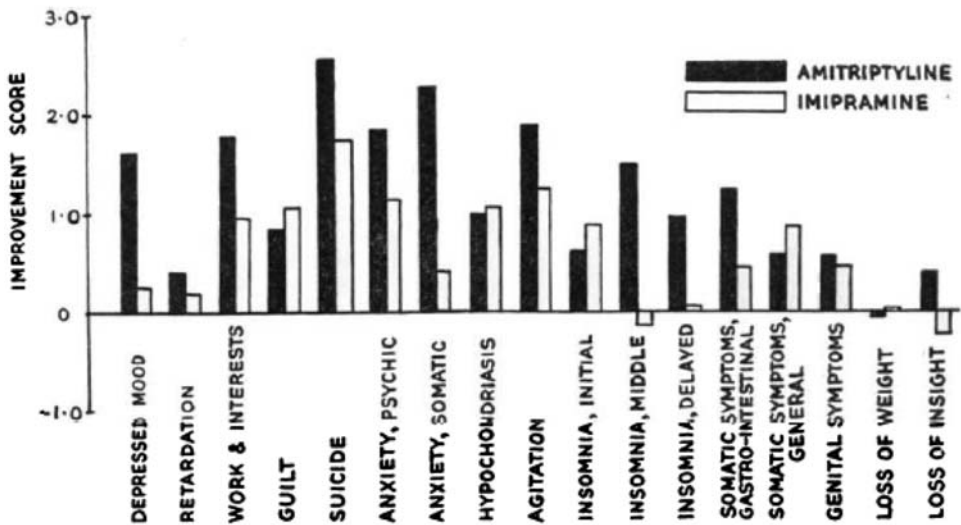


FIG. 2.

Figure 4. Burt CG, et al. Amitriptyline in depressive states: a controlled trial. *Br J Psychiatry* 1962; 108: 711-730.

temporal and experiential dimensions of treatments that were difficult to collect from patients after ECT. Fritz Freyhan, Clinical Director, Director of Research, Delaware State Hospital, Farnhurst, Delaware, explained this point in 1960, showing how drug treatments could be combined with psychotherapy.

The pharmacological treatment of depression offers this immense psychological advantage: the patient maintains his experiential continuity. The amnesic syndrome associated with ECT, to which many attributed therapeutic significance, proves to be quite superfluous as is seen in successful pharmacotherapy. The preservation of experiential continuity has vast implications for psychotherapy. Until now, psychotherapy either followed ECT or had to be limited to patients who seemed capable of affective contact and of self-control over suicidal impulses. With ECT, the patient remains physically and emotionally passive. His recovery comes, as it were, from without. Pharmacotherapy makes him a participating partner. This offers psychotherapy entirely new opportunities to involve the patient in the therapeutic process until recovery is seen as coming from within.²⁰

The second study to use the scale, albeit casually and with crude aggregate scores, was by AA Robin and J Harris at Runwell Hospital, Essex, in a comparison of imipramine and ECT.²¹ In this study, as in many others at this time, ECT was found to give better outcomes.

In 1963, JT Rose published a study of patient responses to ECT using HRSD.²² In measuring the impact of therapy, he validated HRSD by the fact 'that a drop in the score corresponded in the great majority of cases with improvement as recorded by overall clinical assessments and with falling scores in the occupational therapy ratings.' This is interesting as Hamilton developed his scale because of his dissatisfaction with

overall clinical assessments and other scales. Cross reference to, and validation against, overall clinical assessment was common in discussions of HRSD throughout the 1960s and 1970s, not least because the scale was about changing qualitative judgments of clinical outcomes into quantitative values, either in a single score or a matrix of scores.

Interestingly, HRSD was not used in 1964-1965 in a major clinical trial on treatments for depressive illness organised by the Clinical Psychiatry Committee of the Medical Research Council (MRC), even though Hamilton played a leading role in the scheme.²² The trial used both an overall clinical rating of severity and its own scale of 15 symptoms: depressed mood, psychomotor retardation, suicidal ideas, ideas of bodily change, ideas of reference, self-reproach, anxiety, insomnia (early, middle, late) anorexia and fatigue. This scale bore a close relation to HRSD in both the symptoms monitored and the range of scoring, giving tacit endorsement to Hamilton's approach if not his particular scale. In fact, the Committee invented its own so-called 'MRC Scale', which was used quite widely for a number of years, but fell away as HRSD took centre stage.

That the uptake of HRSD was relatively slow is borne out by the number of publications in which it was cited in its first 20 years, see Figure 5, which is presented with all the usual caveats about citations and what they mean. Two sets of data are given: the number of articles each year citing Hamilton's 1960 paper and the number of papers cited with 'depression' in the title. There is steady growth in the number of papers citing HRSD, but this is slower than the overall growth of citations on depression, bearing in mind that both were influenced by the increase in the number of medical journals and the drive to publish more and often. Also, there were many publications, particularly at the end of the

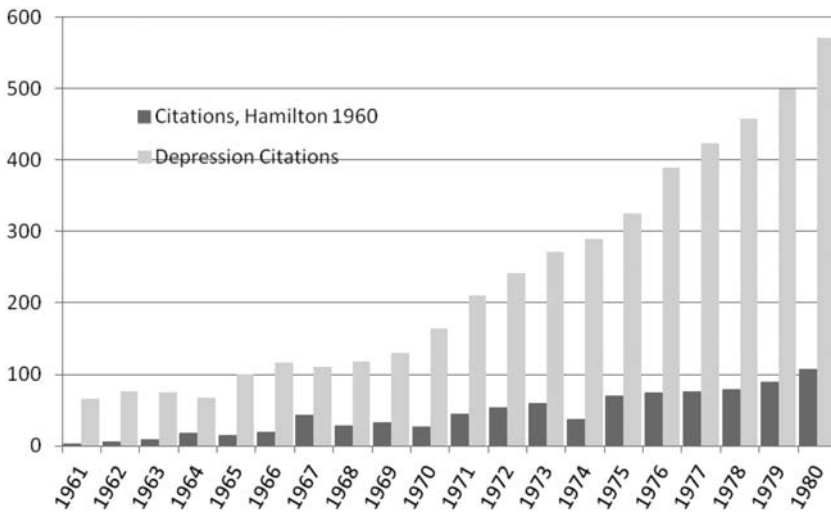


Figure 5. Number of articles each year citing Hamilton M, A rating scale for depression, *J Neurol Neurosurg Psychiatry* 1960; 23: 56–62; and number of articles cited with “depression” in the title. Source: Web of Science.

1970s, in which HRSD was used without citing the 1960 paper. Perhaps it was too well known to need citing? Perhaps the absence of citation indicated that it was being used only casually? And, of course, citation did not mean that authors followed Hamilton’s protocols, in fact psychiatrists used HRSD selectively and flexibly. Writing in 2001, Jane Williams observed that over time,

Several versions of the scale had come into use, with differences in their total number of items, their anchor descriptions, their item interpretations and their scoring conventions By 1990 there were so many versions of the HAM-D that researchers and clinicians had lost track of what was available, and what were the characteristics of each one. No single version of the HAM-D or single set of conventions has been universally accepted.¹¹

Williams noted that by this time, in different publications the number of items scored as HRSD had risen from 17 to 59.¹¹

For much of the 1960s, HRSD was discussed as just another rating scale. For example in 1965, Gerald Klerman and Jonathan Cole’s review of imipramine and related antidepressants mentions HRSD three times in different contexts and always in relation to other scales.²³

Phenomenological differentiations of depressed patients have been developed, using symptom patterns and clusters derived by multivariate statistical techniques. Grinker et al., Friedman et al., Hamilton and Wittenborn et al. have published promising findings.

For example, in studying hospitalized patients, especially severely depressed or schizophrenic patients, well validated scales, particularly by Lorr, Wittenborn, Hamilton and others are widely used. Instruments for nursing observations and for patients’ self-ratings also have been developed.

Drug-placebo differences were revealed by global estimates of degree of depression and by ratings of specific symptoms like

anxiety, insomnia, weight gain and guilt. Hamilton's rating scale, Lorr's Inpatient Multidimensional Psychiatric Scale and the Wittenborn Psychiatric Scale were sensitive to differences in most studies in which they were employed.²³

This illustrates Martin Roth's statement in his brief biography of Max Hamilton that 'It took more than a decade before the HRSD scale was recognised as a major contribution to knowledge and clinical practice.'²⁴

Healy suggests that one reason HRSD was widely used is that it gave particular weight to anxiety symptoms, and thus was good at charting the positive effects of drugs, like imipramine, that were anxiolytic. Alan Broadhurst, a pharmacologist, who was in the group at Geigy that discovered imipramine told David Healy that, 'Max Hamilton was excited about imipramine and it certainly did fit in beautifully with his rating scale. Years later he still referred to it as a happy coincidence'.⁸ However, therapeutic regimes change for so many reasons that it is difficult to tease out the relative importance of HRSD relative to other factors and, although I do not have the data, it is likely that the uptake of imipramine was more rapid than that of HRSD.²⁵

An alternative approach to assessing the rise of HRSD is to look at when and how it was criticised, and why these objections did not impede its progress to becoming the 'Gold Standard.' In the 1960s, HRSD had a competitor, the Inventory for Measuring Depression (then ID and now Beck depression inventory (BDI)), proposed by Aaron T Beck at the University of Pennsylvania.²⁶ BDI has proved similarly enduring and also had the advantage of being a 'first' and the one against which other scales were calibrated and validated. Beck was a pioneer of cognitive therapy and his scale was quite different to HRSD in being based on a patient's self-rating. In its original form the BDI consisted of 21 questions, each with

four possible answers that the patient had to rate 0-3. This gave a theoretical maximum score of 63. A score above 30 indicated severe illness, 19-29 moderate, 10-18 mild and below 10 minimal. A common way of contrasting BID with HRSD was to say that it was 'subjective': it relied upon patients' thoughts and feelings, while HRSD was 'objective', because it was mainly based on clinician observations of bodily and behavioural symptoms.

In 1965, Maryse Metcalfe and Ellen Goldmann compared HRSD favourably with BDI, though they acknowledged that it depended on the skill of the rater and their clinical bias, which, they cautioned, 'made it somewhat difficult to compare meaningfully results obtained in different investigations.'²⁷ In their view, the advantages of BDI were that it was simple, quick and easy to administer, and 'independent of doctors' and nurses' bias, seemingly relying on the 'constant' of the patient. In 1967, John Schwab and colleagues, at the University of Florida College of Medicine, published a comparison of HRSD and BDI amongst ordinary and, one must assume, mostly non-depressed medical inpatients.^{28,29} They found a good correlation ($r_z = 0.75$) in scores, but argued that the two scales were complementary because they measured 'different components of the depressive complex.'

Hamilton assessed and offered a further elaboration of his own scale in 1967.³⁰ The second paper was largely methodological, though it did consider a larger patient group and females as well as males. He also added four extra symptoms to score. However, the article was not easy reading for his peers. It was highly mathematical, as the Abstract illustrates.

'This is an account of further work on a rating scale for depressive states, including a detailed discussion of the general problems of comparing successive samples from

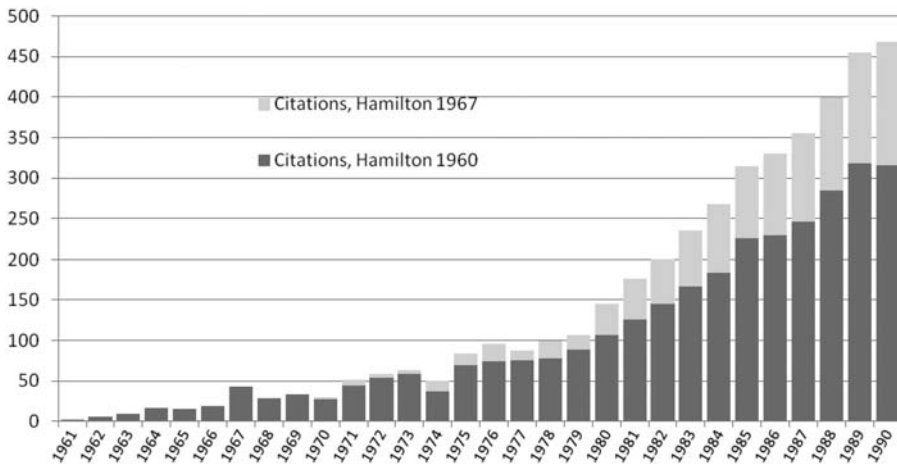


Figure 6. Number of article each year citing Hamilton M, A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; 23: 56–62 and Hamilton M, Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967; 6: 278–296. Source: Web of Science.

a ‘population’, the meaning of the factor scores, and the other results obtained. The intercorrelation matrix of the times of the scale has been factor-analysed by the method of principal components, which were then given a Varimax rotation. Weights are given for calculating factor scores, both for rotated as well as unrotated factors.³⁰

The data to the end of 1990 (Figure 6) shows that, if citations in any way indicate the resources used by psychiatrists in their work, that they stuck with the 1960 paper, for the later elaboration was cited less, even allowing for lags.

In his 1967 paper, Hamilton noted, in a very revealing statement, that this study had been difficult because of the time taken to accumulate a sufficient number of patients with depression. What he actually meant was the difficulty in finding *appropriate* patients, that is, those with treatable illness, as he contrasted this difficulty with the ease of earlier studies with patients in mental hospitals where there were large numbers of chronic cases.³⁰ It seems that within a

decade, what counted as depression, along with who and how they suffered, had changed.

I now want to jump another ten years and consider the ways that HRSD was being used in therapeutic trials at the end of the 1970s.³¹ By this time almost all trials were with psychopharmaceuticals, though ECT was still being used for patients diagnosed with ‘severe’ depression. In fact, prior treatment with ECT often excluded patients from participation in drug trials. However, HRSD was still being used in assessments of ECT, as well as psychotherapy.³² And in 1977, it was even used by Aaron Beck to compare ‘pharmacotherapy’ and ‘cognitive therapy,’ see Figure 7.³³

To sample the uses of HRSD, I surveyed all of the clinical trials for depression published in the medical journals listed in Web of Science for 1979. It was impossible to produce reliable quantitative data of the series, because of the different drugs, protocols and citation practices, so I have chosen to discuss articles that are representative. In most trials HRSD was used with another scale and sometimes with multiple scales, as

Table IV. Clinical Ratings of Severity of Depression at Initiation and Termination of Treatments^a

Time of assessment		Hamilton Rating Scale for Depression			
		Completers only		Completers plus dropouts	
		Cognitive therapy (<i>N</i> = 15)	Pharmacotherapy (<i>N</i> = 14)	Cognitive therapy (<i>N</i> = 16)	Pharmacotherapy (<i>N</i> = 20)
Pretreatment	\bar{X}	21.20	22.43	20.94	21.95
	<i>SD</i>	3.34	4.24	3.40	4.27
Posttreatment	\bar{X}	5.80	9.29	6.25	10.10
	<i>SD</i>	3.67	6.62	3.98	5.94

		Raskin Depression Rating Scale			
		(<i>N</i> = 14)	(<i>N</i> = 10)	(<i>N</i> = 15)	(<i>N</i> = 14)
		Pretreatment	\bar{X}	9.86	10.20
	<i>SD</i>	1.75	.92	1.71	1.41
Posttreatment	\bar{X}	3.93	5.80	4.20	7.10
	<i>SD</i>	1.44	3.49	1.82	3.48

^aReduced *N*s are due to missing pre- or posttreatment scales on some patients.

Figure 7. An example of the reporting outcomes of the use of HRSD with another scale and for different treatments.³³

in the report of a controlled trial of trimipramine and monoamine oxidase inhibitors at St Thomas's Hospital, London, published in 1979. The authors stated:

The patients completed the Beck scale for depression and the Middlesex Hospital Questionnaire (MHQ), and were rated blindly by an independent assessor on the Hamilton rating scale for depression, the MRC depression scales, and an overall six-point rating of the severity of depression. A standard rating of side effects was completed by the psychiatrist who regulated drug dosage to prevent knowledge of any such effects biasing the clinical ratings of the other assessor.^{34,35}

The graphs below show how the results of the different scales were mapped for the six weeks of the trial (Figure 8).

The same pattern was evident in a study of Limbitrol in California.

The patients were evaluated at baseline using the Hamilton rating scale for primary depressive illness (HDS) and the Covi anxiety scale. In addition, the patients completed the short form of the BDI and the Hopkins symptom checklist (SCL-58). Efficacy was assessed at follow-up visits after 1, 2, and 4 weeks of treatment by the physician, using the HDS and a global evaluation, and by the patient using the BDI, the SCL-58, and a global evaluation. In most instances, the BDI and the SCL-5g were completed by the patient prior to his seeing the psychiatrist.³⁶

In a trial of Lithium, HRSD was set against a 5-point nurse rating scale (Figure 9).³⁷

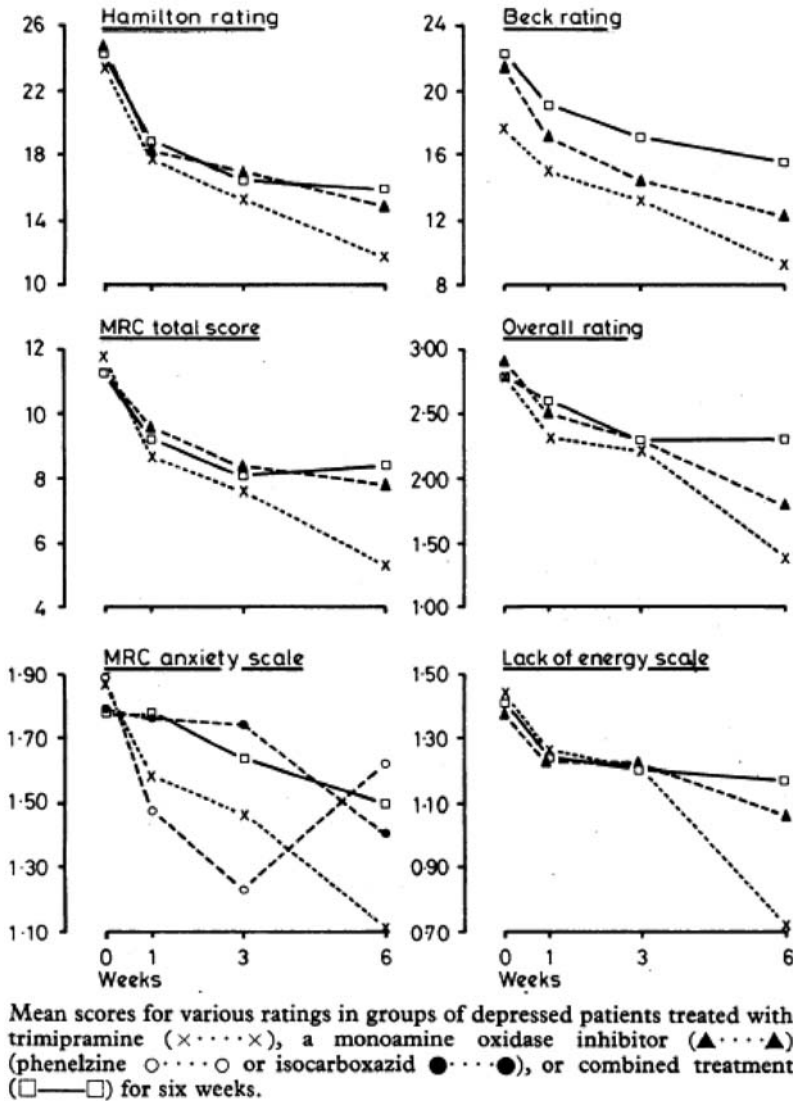


Figure 8. An example of HRSD scores reported against many other scales.³⁴

There are very few publications where the score was disaggregated and the different components mapped to identify specific changes, one exception was a study comparing amineptine and amitriptyline at Hôpital de St. Germain-en-Laye.³⁸ The changes in the total scores were first presented (Figure 10) and when the component scores were set out it was difficult to see

the wood for the trees (Figure 11), and then only 14 out of the 26 items scored had statistical significance.

Discussion

In this paper, I have made two main claims, first that HRSD was applied by clinicians to construct depression as a time-limited

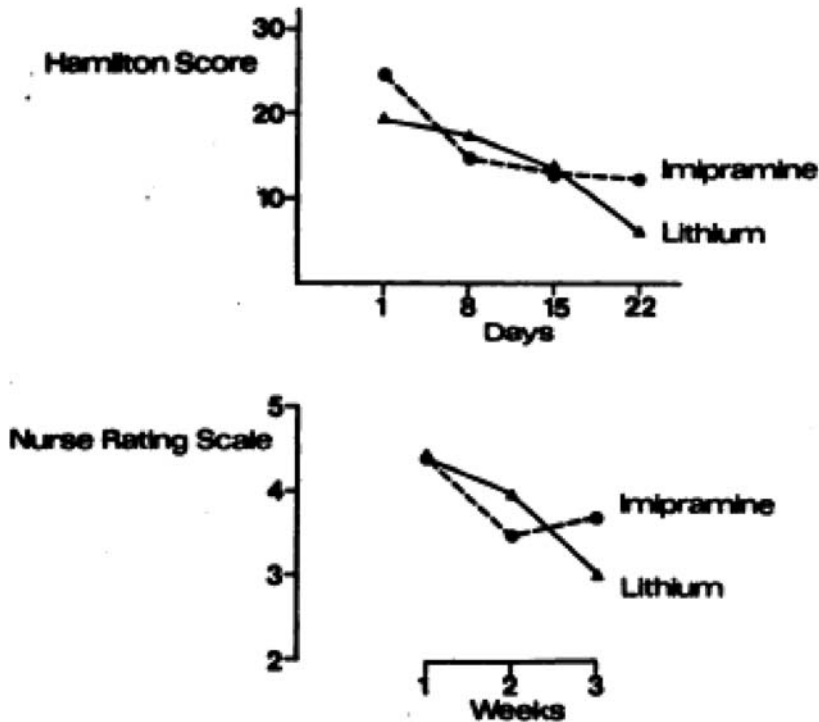


FIG 1.—Trial I-L. Mean Hamilton and nurse ratings.

Figure 9. An example of reporting HRSD in comparison with a nurse rating scale.³⁷

illness, and second, that this influential framing of the condition was used alongside other scales and only rose to dominance gradually. The assumption of the time-limited illness supports the claim of Healy and others that an HRSD-structured characterisation of depression was suited to drug therapy and the interests of pharmaceutical companies in the 1960s and 1970s. The view of psychiatrists in the first half of the twentieth century was that depressive mental illness was chronic, either because of patient susceptibilities rooted in somatic factors, such as hereditary or physical disease, or in psychic variables influenced by upbringing, interpersonal relationships or personality. There was however some turnover in mental hospital patients and moves

to treat many sufferers as out-patients. The patient population peaked in Britain in 1954 at 140,000, when there were 121,000 beds, suggesting that turnover was not great and that most patients had chronic conditions. The rundown in the number of beds and the move to community care saw depression move out of the hospital and into the community, as an out-patient or general practitioner managed condition. In this setting, and due to new framings and new treatments, it was approached as a 'mild' and short-lived condition, at least compared to the illness that had previously required hospitalisation.³⁹ HRSD was used to frame this new 'depression' and its sufferers, normalising it to the ways of seeing and treating illness as a treatable episode or

Table III. Mean (\pm S.D.) total scores on Hamilton Rating Scale before and during treatment

Assessment	Amineptine	Amitriptyline
Before treatment	50.92 \pm 1.18	52.12 \pm 1.42
After 15 days	49.24 \pm 1.18**	50.80 \pm 1.55*
After 30 days	41.62 \pm 1.12***	43.00 \pm 1.43***
After 45 days	36.54 \pm 1.25***	38.22 \pm 1.48***

Significance of difference from before treatment * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ or less

Figure 10. A typical use of HRSD charting the effects of two drugs over time.³⁸

Table IV. Changes in mean (\pm S.D.) scores for individual items of Hamilton Rating Scale during treatment

Item	Drug	Before treatment	After treatment		
			Day 15	Day 30	Day 45
1. Depressed mood	Amineptine	3.20 \pm 0.11	3.08 \pm 0.09*	2.41 \pm 0.11***	1.79 \pm 0.13***
	Amitriptyline	3.20 \pm 0.10	3.20 \pm 0.13	2.50 \pm 0.14***	1.94 \pm 0.17***
2. Guilt feelings	Amineptine	1.60 \pm 0.14	1.60 \pm 0.14	1.33 \pm 0.09***	1.20 \pm 0.08***
	Amitriptyline	1.96 \pm 0.14	1.92 \pm 0.14	1.40 \pm 0.11***	1.22 \pm 0.10***
3. Suicide	Amineptine	1.76 \pm 0.15	1.64 \pm 0.16	1.29 \pm 0.11***	1.04 \pm 0.04***
	Amitriptyline	1.84 \pm 0.13	1.76 \pm 0.16	1.20 \pm 0.09***	1.11 \pm 0.08***
4. Initial insomnia	Amineptine	2.20 \pm 0.12	2.20 \pm 0.12	1.91 \pm 0.13***	1.75 \pm 0.12***
	Amitriptyline	2.24 \pm 0.12	2.16 \pm 0.11	1.90 \pm 0.10**	1.89 \pm 0.11**
5. Middle insomnia	Amineptine	1.60 \pm 0.10	1.60 \pm 0.10	1.50 \pm 0.10	1.25 \pm 0.09***
	Amitriptyline	1.84 \pm 0.12	1.80 \pm 0.12	1.45 \pm 0.14***	1.22 \pm 0.10***
6. Delayed insomnia (early morning waking)	Amineptine	1.32 \pm 0.09	1.24 \pm 0.08	1.12 \pm 0.06	1.00 \pm 0.01***
	Amitriptyline	1.44 \pm 0.13	1.32 \pm 0.11	1.25 \pm 0.12	1.11 \pm 0.08***
7. Work and activity	Amineptine	3.40 \pm 0.14	3.08 \pm 0.17***	2.58 \pm 0.15***	2.16 \pm 0.16***
	Amitriptyline	3.44 \pm 0.19	3.28 \pm 0.17**	2.50 \pm 0.17***	2.16 \pm 0.23***
8. Retardation	Amineptine	1.96 \pm 0.14	1.92 \pm 0.11	1.58 \pm 0.13***	1.20 \pm 0.08***
	Amitriptyline	2.00 \pm 0.15	2.00 \pm 0.15	1.60 \pm 0.16	1.27 \pm 0.15***
9. Agitation	Amineptine	1.64 \pm 0.11	1.76 \pm 0.10	1.41 \pm 0.10	1.16 \pm 0.07***
	Amitriptyline	1.68 \pm 0.12	1.72 \pm 0.13	1.30 \pm 0.10***	1.11 \pm 0.07***
10. Psychic anxiety	Amineptine	3.36 \pm 0.14	3.16 \pm 0.16**	2.37 \pm 0.13***	1.70 \pm 0.11***
	Amitriptyline	3.20 \pm 0.14	3.04 \pm 0.17	2.35 \pm 0.16***	2.00 \pm 0.11***
11. Somatic anxiety	Amineptine	2.36 \pm 0.15	2.36 \pm 0.15	1.91 \pm 0.11***	1.79 \pm 0.12***
	Amitriptyline	2.28 \pm 0.17	2.28 \pm 0.16	2.10 \pm 0.14**	1.94 \pm 0.15***
12. Gastro-intestinal symptoms	Amineptine	1.68 \pm 0.09	1.68 \pm 0.09	1.45 \pm 0.10**	1.33 \pm 0.09***
	Amitriptyline	1.68 \pm 0.13	1.68 \pm 0.13	1.65 \pm 0.15	1.50 \pm 0.16**
13. General somatic symptoms	Amineptine	1.52 \pm 0.10	1.52 \pm 0.10	1.45 \pm 0.10	1.29 \pm 0.09***
	Amitriptyline	1.64 \pm 0.09	1.64 \pm 0.09	1.50 \pm 0.11	1.33 \pm 0.11***
14. Genital symptoms	Amineptine	2.80 \pm 0.21	2.80 \pm 0.21	2.79 \pm 0.23	2.58 \pm 0.24*
	Amitriptyline	2.80 \pm 0.19	2.76 \pm 0.20	2.90 \pm 0.17	2.83 \pm 0.23

Figure 11. Reporting disaggregated HRSD scores, as illustrated above, became less common.³⁸

episodes, rather than a life course condition. As such, HRSD served the interests of psychiatrists and psychiatry in the new era of treating specific illnesses outside of mental hospitals.

HRSD rose to dominance ‘from below.’ When it was sanctioned ‘from above’ in the 1980s, by the World Health Organisation, Food and Drugs Administration, and other

medicine licensing agencies, this was acknowledging its widespread use, not creating it ‘top down.’ Paradoxically, the eventual dominance of HRSD was in large part due to its successful validation against the holistic clinician assessments, the very thing Hamilton designed it against. However, HRSD was a clinician scoring instrument and proved simple to use because clinicians

made it so, choosing overall scores rather than disaggregated or factor scores. In many ways, the 'S' in HRSD stood for 'Score' not 'Scale', but either way it was a quantitative datum on a relatively large and finely grained scale of 100, at least when compared to the previous clinician scales. Overall, HRSD was a strange kind of 'standard,' being quite non-standard in the flexible and widely varying ways it was used, the number and type of items in the scale and the meanings given to its findings.

Acknowledgments

I would like to thank my colleagues Carsten Timmermann, Robert GW Kirk, Stephanie Snow, David Thompson and Duncan Wilson for comments on earlier versions of the paper and the referee who suggested that I say more about DSM-II and -III. This work was first presented to the ESF-funded Drug Standards, Standard Drugs meeting on 'The view from below: On standards in clinical practice and clinical research,' at the Charité Institute for the History of Medicine, Berlin, in September 2012.

Funding

This work was supported by the Wellcome Trust (Grant Number 092782).

References

- Andrews G. Should depression be managed as a chronic disease? *Br Med J* 2001; 1: 419–421.
- Gask L. Is depression a chronic illness? For the motion. *Chronic Illn* 2005; 1: 101.
- Monroe SM and Harkness KM. Editorial: is depression a chronic mental illness? *Psychol Med* 2012; 42: 899–902.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; 23: 56–62.
- In Conversation with Max Hamilton. *Psychiatr Bull* 1983; 7: 62–66.
- Bagby RM. The Hamilton Depression Rating Scale: has the Gold Standard become a lead weight? *Am J Psychiatry* 2004; 161: 2163–2177.
- Greenberg G. *Manufacturing depression: the secret history of a modern disease*. London: Bloomsbury, 2010.
- Healy D (ed.) *The psychopharmacologists*. London: Altman, 1996.
- Tansey T. Review of D Healy, 'The Antidepressant Era'. *Hist Psychiatry* 1998; 9: 536.
- Healy D. *The creation of psychopharmacology*. Cambridge, MA: Harvard University Press, 2002, p.350.
- Williams JBW. Standardizing the Hamilton Depression Rating Scale: past, present, and future. *Eur Arch Psychiatry Clin Neurosci* 2001; 251(Suppl. 2): II/6–II/12.
- Mayes R and Horowitz AV. DSM-II and the revolution in the classification of mental illness. *J Hist Behav Sci* 2005; 41: 249–267.
- Wilson M. DSM-III and the transformation of American psychiatry: a history. *Am J Psychiatry* 1993; 150: 399–410.
- Rollin RH. Hamilton, Max (1912–1988). *Oxford Dictionary of National Biography*. Oxford University Press, online ed., September 2010, <http://www.oxforddnb.com/view/article/70380> (2004, accessed 16 July 2012).
- Hargreaves GR, Hamilton M and Roberts JM. Bentactzine as an aid in the treatment of anxiety states. *Br Med J* 1957; 1: 306–310.
- Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959; 32: 50–55.
- Hamilton M and White JM. Clinical syndromes in depressive states. *Br J Psychiatry* 1959; 105: 985–998.
- Hamilton M and White JM. Factors related to the outcome of depression treated with ECT. *Br J Psychiatry* 1960; 106: 1031–1041.
- Burt CG, et al. Amitriptyline in depressive states: a controlled trial. *Br J Psychiatry* 1962; 108: 711–730.
- Freyhan F. The modern treatment of depressive disorders. *Am J Psychiatry* 1960; 116: 1057–1064.
- Robin AA and Harris JA. A controlled comparison of imipramine and electroplexy. *Br J Psychiatry* 1962; 108: 217–219.
- Rose JT. Reactive and endogenous depressions – response to ECT. *Br J Psychiatry* 1963; 109: 213–217.
- Klerman G and Cole J. Clinical pharmacology of imipramine and related antidepressant compounds. *Pharmacol Rev* 1965; 17: 101–141.
- Roth M. Max Hamilton: a life devoted to psychiatry. In: Bech P, Coppen A (eds) *The Hamilton Scales*. Berlin: Springer-Verlag, 1990, pp.1–9.
- Tansey EM. They used to call it psychiatry: aspects of the development and impact of

- psychopharmacology. In: Gisjwijt-Hofstra M and Porter R (eds) *Cultures of psychiatry and mental health care in postwar Britain and the Netherlands*. Amsterdam: Rodopi, S. 81, pp.79–102.
26. Beck AT, et al. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; 4: 561–571.
 27. Metcalfe M and Goldman E. Validation of an inventory for measuring depression. *Br J Psychiatry* 1965; 111: 240.
 28. Schwab JJ, et al. A comparison of two rating scales for depression. *J Clin Psychiatry* 1967; 23: 94–96.
 29. Schwab JJ, et al. Hamilton Rating Scale for Depression with medical inpatients. *Br J Psychiatry* 1967; 113: 83–88.
 30. Hamilton M. Development of a rating scale for primary depressive illness. *British J Soc Clin Psychol* 1967; 6: 278–296.
 31. Hughes JR, et al. Measurement of depression in clinical trials: an overview. *J Clin Psychiatry* 1982; 43(3): 85–88.
 32. Lambourn J and Gill D. A controlled comparison of simulated and real ECT. *Br J Psychiatry* 1978; 133: 514–519.
 33. Rush AJ, et al. Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognit Ther Res* 1977; 1: 17–37.
 34. DiMascio A, et al. Differential symptom reduction by drugs and psychotherapy in acute depression. *Arch Gen Psychiatry* 1979; 36: 1450–1456.
 35. Young JPR, et al. Controlled trial of trimipramine, monoamine oxidase inhibitors, and combined treatment in depressed outpatients. *Br Med J* 1979; ii: 1315–1317.
 36. Feighner JP, et al. A placebo-controlled multicenter trial of limbitrol versus its components (amitriptyline and chlordiazepoxide) in the symptomatic treatment of depressive illness. *Psychopharmacology* 1979; 61: 217–225.
 37. Worrall EP, et al. Controlled studies of the acute antidepressant effects of lithium. *Br J Psychiatry* 1979; 135: 255–262.
 38. van Amerongen P. Double-blind clinical trial of the antidepressant action of amineptine. *Curr Med Res Opin* 1979; 6: 93–101.
 39. Mitchell-Heggs N. Aspects of the natural history and clinical presentation of depression. *Proc R Soc Med* 1971; 64: 1171–1174.