# ROUTLEDGE HANDBOOK OF PHYSICAL ACTIVITY AND MENTAL HEALTH

EDITOR-IN-CHIEF
*Panteleimon Ekkekakis*

SECTION EDITORS
*Dane B. Cook*
*Lynette L. Craft*
*S. Nicole Culos-Reed*
*Panteleimon Ekkekakis*
*Jennifer L. Etnier*
*Mark Hamer*
*Kathleen A. Martin Ginis*
*Justy Reed*
*Jasper A.J. Smits*
*Michael Ussher*

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2013
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2013 Panteleimon Ekkekakis

The right of the editor to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*Trademark notice*: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

*British Library Cataloguing in Publication Data*
A catalogue record for this book is available from the British Library

*Library of Congress Cataloging in Publication Data*
[CIP data]

ISBN: 978–0–415–78299–9 (hbk)
ISBN: 978–0–203–13267–8 (ebk)

Typeset in Bembo
by Keystroke, Station Road, Codsall, Wolverhampton

# PHYSICAL ACTIVITY AS A MENTAL HEALTH INTERVENTION IN THE ERA OF MANAGED CARE

## A rationale

*Panteleimon Ekkekakis*

The aim of this introductory chapter is to outline the *raison d'être* of this handbook. This is established by a sequence of arguments substantiating the following points. First, mental health problems are very prevalent in the industrialized world, are of exceptional societal importance, and exact a high cost both in terms of quality of life and in terms of economic impact. Second, a high percentage of individuals suffering from mental health problems do not seek help, and of those who do, many exhibit poor compliance with treatment regimens. Most likely, these phenomena are due to concerns related to the high cost of diagnosis and treatment, questions about the effectiveness of common treatments, and an unwillingness to risk or cope with adverse side-effects. Third, a growing corpus of evidence suggests that physical activity can be an inexpensive, safe, and, most importantly, effective lifestyle intervention that could be used to both prevent and treat a wide range of mental health problems and to improve overall quality of life. However, the effectiveness of physical activity as a mental health intervention is not universally acknowledged. For various reasons, many researchers and practitioners in medicine, psychology, and public health remain either uninformed or unconvinced about the potential of physical activity to promote mental health. Some of these reasons may stem from dualistic and disciplinary rifts and others from legitimate concerns about the methodological rigor of the extant studies. Distinguishing between these, and arriving at an accurate appraisal of the evidence, is becoming increasingly difficult.

This volume was compiled with the goal of offering an encompassing and balanced account of the evidence. World-leading experts from various disciplines review what has been found so far, highlighting notable accomplishments but also pointing out limitations and persistent lacunae that should be addressed in future studies.

### Societal importance of public mental health

The prevalence, societal impact, and economic cost of mental disorders are impossible to estimate with accuracy. It is generally acknowledged that published figures most likely represent underestimates. To a large extent, this is attributed to the stigma still attached to mental disorders and the associated unwillingness of many people to undergo a diagnosis for a mental disorder or to disclose information that they perceive as personally embarrassing.

## *Worldwide impact*

The World Health Organization (2003) estimates that approximately 450 million people suffer from a mental or behavioral disorder. Worldwide, four of the six leading causes of disability are psychiatric in nature (i.e., depression, alcohol-use disorders, schizophrenia, and bipolar disorder). Collectively, mental disorders (excluding those associated with stroke and brain trauma) account for 12.9% of all disability-adjusted life years lost, compared to 9.3% for cardiovascular disease and 5.2% for all cancers combined (Andlin-Sobocki, Jönsson, Wittchen, & Olesen, 2005). As the prevalence and impact of mental health problems grow, the 14-country World Mental Health Survey shows that 35.5–50.3% of serious cases of mental illness in developed countries and 76.3–85.4% in less-developed countries receive no treatment (WHO World Mental Health Survey Consortium, 2004). Besides social stigmatization, the high cost of therapy and the scarcity of specialized therapeutic resources are believed to contribute to this problem.

In 2000, depression was the fourth leading cause of disease burden according to the Global Burden of Disease (GBD) study, accounting for 4.4% of total disability-adjusted life years. It was also responsible for the largest portion of non-fatal burden, accounting for almost 12% of all years lived with disability (Üstün, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004). It is estimated that by the year 2020, unipolar major depression will become the second cause of disability worldwide, behind only ischemic heart disease (Murray & Lopez, 1997). By 2030, unipolar depression is projected to be the leading cause of disability in high-income countries (Mathers & Loncar, 2006).

Moreover, dementia, one of the most debilitating mental disorders, is estimated to afflict more than 30 million people worldwide (Ferri et al., 2005; Wimo, Jönsson, & Winblad, 2006). With many countries facing the challenge of an aging population, 4.6 million new cases of dementia are expected every year (one every 7 seconds). At this rate, the number of people affected by this disease is expected to double every 20 years, exceeding 80 million by 2040. In 2009, the total worldwide societal cost of dementia, based on a population of 34.4 million demented persons, was estimated at US$422 billion, including US$142 billion (34%) for informal care. This estimate represents an increase of 34% between 2005 and 2009 (Wimo, Winblad, & Jönsson, 2010).

Across the 27 countries/members of the Europe Union, it is estimated that each year 38.2% of the population suffers from a mental disorder (Wittchen et al., 2011). Even after accounting for comorbidity, this figure corresponds to 164.8 million people. Brain and mental disorders collectively account for 26.6% of the total disability burden. The four most disabling conditions are depression, dementias, alcohol use disorders, and stroke. Very importantly, fewer than one-third of the people affected receive any treatment.

In the United States, according to the replication of the National Comorbidity Survey, about half of Americans meet the criteria for a DSM-IV disorder sometime in their life. Approximately 50% of all lifetime cases of mental illness begin by age 14 and 75% by age 24 (Kessler et al., 2005). Importantly, there are long delays – sometimes decades – between the first onset of symptoms and when people seek treatment (Wang et al., 2005). According to the report by the U.S. Surgeon General on mental health (United States Department of Health and Human Services, 1999), "concerns about the cost of care – concerns made worse by the disparity in insurance coverage for mental disorders in contrast to other illnesses – are among the foremost reasons why people do not seek needed mental health care" (p. 23). Moreover, there are considerable disparities along socioeconomic, geographical, and racial/ethnic lines (González et al., 2010). As a result, according to the data from the National Comorbidity Survey, a nationwide prevalence study based on face-to-face household interviews, only 20.3% of individuals who satisfied diagnostic criteria for a mental disorder received treatment in 1990–1992 and 32.9% in 2001–2003

(Kessler et al., 2005). Interestingly, of those individuals who received treatment, only half met diagnostic criteria for a mental disorder.

## Economic cost

The economic cost of mental illness at the societal level is both direct and indirect, the latter being considerably more difficult to estimate. Direct costs include expenses associated with medication use, clinic visits, and hospitalization. The Agency for Healthcare Research and Quality reported that the number of individuals treated for mental disorders nearly doubled from 19.3 million in 1996 to 36.2 million in 2006 (Soni, 2009). The direct cost of their care increased from US$35.2 billion to US$57.5 billion. This made mental disorders the third costliest in terms of direct costs, behind only heart disease (US$78.0 billion) and trauma-related disorders (US$68.1 billion) and at the same level as cancer (US$57.5 billion). However, this estimate does not include several major categories of direct costs, such as drugs. A study conducted by the Substance Abuse and Mental Health Services Administration (Mark et al., 2007) estimated the direct cost of mental health disorders at US$100 billion in 2003, with substance abuse problems costing an additional US$21 billion. These figures exclude costs associated with all types of dementia, tobacco addiction, mental retardation, and mental developmental delays. An estimate subject to the same exclusions puts the annual cost at US$203 billion by 2014 (Levit et al., 2008).

The indirect costs stem from a multitude of sources, including (but not limited to) reduced productivity and increased absenteeism, reduced income and ensuant need for publicly financed assistance (e.g., unemployment compensation, disability benefits, food stamps, public housing), reduced educational attainment, incarceration and other crime-related costs, and homelessness. According to one study (Kessler et al., 2008), after controlling for confounding variables (e.g., age, gender, race, region of the country, as well as alcohol and illicit drug abuse), an individual who had experienced a mental illness in the previous 12 months had an income that was on average US$16,306 less than that of an individual without a mental illness (US$26,435 among men, US$9,302 among women). At the societal level, the cost was astronomical, estimated at US$193.2 billion (US$131.3 billion among men, US$61.9 billion among women). What complicates the estimation of the total societal cost further is the fact that mental disorders exhibit extensive comorbidity with a broad range of physical illnesses, forming a complex web of causal interactions. For example, persons with serious mental illness smoke 44% of all cigarettes in the United States, thus increasing the prevalence of pulmonary disease (Insel, 2008). A study across eight U.S. states documented that clients of public mental health agencies had a risk of death that was 1.2 to 4.9 times higher than that of the average state population. The lives of these individuals were 13.5 to 32.2 years shorter than state averages (Colton & Manderscheid, 2006).

To illustrate the grave economic implications of indirect costs, it is useful to examine two disorders, the impact of which has been analyzed in numerous studies. In 2000, the cost of depression in the United States was estimated at US$83.1 billion. Of this amount, US$26.1 billion (31%) were direct medical costs and US$5.4 billion (7%) were suicide-related mortality costs. On the other hand, the workplace costs were estimated at US$51.5 billion or 62% of the total (Greenberg et al., 2003).

According to the Alzheimer's Association (2010), Alzheimer's disease (AD) is the seventh leading cause of death in the United States, causing approximately as many deaths annually as diabetes. AD is also the fifth leading cause of death among Americans aged 65 and older. Importantly, although the rates of other major causes of death have declined in recent years, the rate due to AD has increased and is projected to continue to increase. For example, between 2000 and 2006, heart disease deaths decreased by 11.1%, stroke deaths decreased by 18.2%, and

prostate cancer-related deaths decreased by 8.7%. In contrast, deaths due to AD increased by 46.1%. By 2050, the prevalence of AD in the United States is expected to reach 11 to 16 million people, as the baby boomer generation ages. The Alzheimer's Association estimates that the direct cost of caring for people over 65 with AD and other dementias will reach US$172 billion in 2010, including US$123 billion for Medicare and Medicaid. In 2004, each Medicare beneficiary with AD or another dementia incurred costs (for health care, long-term and hospice) that were three times higher than those of other Medicare beneficiaries of the same age. Likewise, in 2004, each Medicaid beneficiary with AD or another dementia incurred costs (for nursing home and other long-term care for people with very low income and low assets) that were nine times higher than those of other Medicaid beneficiaries of the same age. In addition to these costs, in 2009, approximately 11 million family members and other unpaid caregivers provided an estimated 12.5 billion hours of care to persons with AD and other dementias. This care is valued at nearly US$144 billion. Moreover, caregivers of individuals suffering from AD and other dementias experience increased risk of anxiety, depression, and other mental health problems (Cooper, Balamurali, & Livingston, 2007; Cuijpers, 2005; Etters, Goodall, & Harrison, 2008; Schulz, Boerner, Shear, Zhang, & Gitlin, 2006).

In Canada, Health Canada (2002) estimates that 20% of Canadians will experience a mental illness during their lifetimes. The cost was initially estimated at CA$14.4 billion in 1998 (Stephens & Joubert, 2001). The Canadian Mental Health Association (2004) reported to the House of Commons that mental illnesses account for 50% of physicians' billing and is responsible for more hospital bed days than cancer. After taking into account the costs associated with undiagnosed mental illnesses and the impact on health-related quality of life, a more recent estimate raised the cost for 2003 to CA$51 billion (Lim, Jacobs, Ohinmaa, Schopflocher, & Dewa, 2008).

In England, an extensive study by the King's Fund estimated that 8.65 million people with a mental health problem lived in the country in 2007 (McCrone, Dhanasiri, Patel, Knapp, & Lawton-Smith, 2008). A 14% increase is projected by 2026. The associated costs are expected to increase from £48.6 billion in 2007 to £88.45 billion in 2026. In 2007, almost half of the cost (£22.5 billion) represented money spent on direct NHS and social care services to support people with mental disorders. The rest (£26.1 billion) represented the estimated cost to the economy of earnings lost because of the thousands of people unable to work due to their mental illness. The direct (mental health service) costs were projected to increase by 45% over the next 20 years from £22.5 billion to £47 billion due to more cases of dementia and the projected increase in health care costs. Specifically, due to the aging population, the incidence of dementia is expected to increase by 61%, from 582,827 cases to 937,636 cases. The service costs associated with dementia are far higher than all other conditions put together. In 2007, they accounted for 66% of all mental health service costs, but by 2026 it is estimated that they will make up 73% of all mental health service costs.

## Growing skepticism about psycho- and pharmacotherapies

The "formal" methods of treating mental health problems, according to the U.S. Surgeon General's report on mental health (1999), include various types of psychotherapy and pharmacotherapy. Of the two approaches, the use of psychotherapy appears to be steady or declining, whereas the use of pharmacotherapy has increased dramatically in the past several years. The reasons include the scarcity of psychotherapy resources, the relative affordability of pharmacotherapy for most individuals with insurance coverage, and the very aggressive direct-to-consumers advertising of major classes of psychotropic medications (e.g., antidepressants, anxiolytics, mood stabilizers, analgesics, sleep aids).

## Psychotherapy

In the United States, the rates of psychotherapy appear to have remained steady, with 3.2% of individuals in 1987, 3.6% in 1997, and 3.18% in 2007 receiving some type of psychotherapy on an outpatient basis (Olfson & Marcus, 2010; Olfson, Marcus, Druss, & Pincus, 2002). In particular, psychotherapy delivered by a psychiatrist declined from 44.4% of all psychotherapy visits in 1996–1997 to 28.9% in 2004–2005 (Mojtabai & Olfson, 2008). Among many academics, the effectiveness of psychotherapy has been in question since Eysenck's (1952) challenge to the field more than 60 years ago. Eysenck argued that the studies that were available until that time "failed to prove that psychotherapy, Freudian or otherwise, facilitates the recovery of neurotic patients. They show that roughly two-thirds of a group of neurotic patients will recover or improve to a marked extent within about two years of the onset of their illness, whether they are treated by means of psychotherapy or not" (p. 322). He added that "definite proof would require a special investigation, carefully planned and methodologically more adequate than these *ad hoc* comparisons" (p. 323). Even after so many years, this kind of "definite proof" remains elusive (Nutt & Sharpe, 2008).

A persistent source of controversy is the fact that meta-analyses generally show no difference between the effect sizes associated with psychotherapy and those associated with placebos. For example, Prioleau, Murdock and Brody (1983) examined 32 studies comparing psychotherapy to a placebo condition and found an average effect size of only .15. Although some subsequent meta-analysts have questioned this finding (Grissom, 1996; Lipsey & Wilson, 1993), a more recent analysis by Baskin, Callen Tierney, Minami, and Wampold (2003) concluded that as long as the placebo interventions are properly designed (i.e., controlling for such structural aspects as number and duration of sessions, training of therapist, format of therapy, etc.), the effect size compared to psychotherapy is indeed very small (.15). The skepticism is fueled by consistent findings that (a) a substantial percentage of patients report an early positive response (Lambert, 2005) and (b) the various types of psychotherapy, despite large differences in their theoretical background and the format of their procedures, seem indistinguishable in terms of therapeutic effectiveness (Smith & Glass, 1977; Wampold et al., 1997). In response to such findings, several psychotherapists have argued that the placebo is not an appropriate control or comparison condition for assessing the effectiveness of psychotherapy (Herbert & Gaudiano, 2005; Kirsch, 2005; Parloff, 1986; Senger, 1987). The main reason in support of this argument is that what is implied by a "placebo" response is in fact an inextricably intertwined component of the psychotherapeutic process. In essence, the argument is that the so-called "common factors" of all psychotherapeutic modalities (e.g., expectations of benefit, prescription for activity, discussion, attention, compassion, caring, and warmth) may be crucial contributors to the effectiveness of psychotherapy. What this implies, however, is that the therapeutic "active ingredient" in psychotherapy (much less the specific "active ingredient" of different types of psychotherapy) remains unknown. According to a recent authoritative review, "it is remarkable that after decades of psychotherapy research, we cannot provide an evidence-based explanation for how or why even our most well studied interventions produce change" (Kazdin, 2007, p. 23). In other words, if psychotherapy works, no one yet knows why. Moreover, concerns about the safety of psychotherapy, particularly long-term psychotherapy, are growing (Barlow, 2010; Nutt & Sharpe, 2008).

## Pharmacotherapy

The effectiveness and safety of certain types of pharmacotherapy for mental disorders is becoming the subject of intense controversy. In particular, a heated debate has centered around antidepressants, which, since 2004–2005, have become the most widely prescribed class of drugs in

the United States (United States National Center for Health Statistics, 2010). Because of their overwhelming popularity, antidepressants will be used here as an example to illustrate the argument about the growing concerns surrounding the effectiveness and safety of pharmacotherapy for mental disorders.

Across the 27 countries/members of the European Union, 8% of the population overall, and 10% of middle-aged adults, take antidepressants each year (Blanchflower & Oswald, 2011). Antidepressant use exhibits an inverted U-shape relationship with age, being highest for individuals in their late 40s (particularly women) and those who are unemployed, poorly educated, and divorced or separated.

The prescription of antidepressants in the United States nearly doubled in a decade (from 5.84% in 1996 to 10.12% in 2005, or from 13.3 to 27.0 million users; Olfson & Marcus, 2009). Interestingly, during this period, only 26% of the individuals who received prescriptions for antidepressants were being treated for depression, since these drugs were also commonly prescribed for a wide range of mental health conditions, including back pain and headaches, anxiety, adjustment disorders, sleep disorders, neuropathy, fatigue, and bipolar disorder.

In Canada, antidepressants are the most widely prescribed class of psychotropic medications (Beck et al., 2005), with 5.8% of the population using them. Interestingly, nearly half (47.7%) of antidepressant users have no history of depression (Beck et al., 2005). Antidepressants are even prescribed more frequently than sedatives or hypnotics for individuals with anxiety and more frequently than mood stabilizers for individuals with bipolar disorder. Older tricyclic antidepressants, in particular, are now being prescribed predominantly for diagnoses other than depression (Patten, Esposito, & Carter, 2007). At the same time, 59.6% of individuals with a depression diagnosis in the previous year do not receive a prescription for antidepressants.

In Scotland, the volume of prescribed antidepressants tripled between 1995/1996 and 2006/2007 (Lockhart & Guthrie, 2011). This dramatic increase has been due to a combination of more patients being given prescriptions (from 8.0% of the population in 1995/1996 to 11.9% in 2000/2001) and increased treatment duration (from 170 days per year to 200). Across the United Kingdom, the number of prescriptions per patient per year increased from 2.8 in 1993 to 5.6 in 2004 (Moore et al., 2009). In England, antidepressant prescriptions increased by 10% per year, on average, over a 13-year period (i.e., 130% overall; Ilyas & Moncrieff, 2012).

The increasing use of antidepressants does not only have economic implications. Two additional problems are receiving increasing attention in the literature. The first deals with questions about their actual effectiveness and the second with their safety profile. On both of these issues, the initial assumptions of clinicians and the positive impressions among the general public must be reevaluated on the basis of emerging evidence.

Meta-analyses on the effectiveness of antidepressants had reached positive conclusions but also included caveats about the fact that the effect sizes were typically small, the methodological quality typically low, the treatment periods short (6–8 weeks), and pharmaceutical companies were providing the funding (Arroll et al., 2009; Arroll et al., 2005). Although the impact of the strong grip of the pharmaceutical industry on the publication of evidence had long been suspected (Healy & Cattell, 2003; McGoey, 2010), the magnitude of this problem was unveiled in a landmark study published in the prestigious *New England Journal of Medicine* (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008). The researchers compared data from 74 randomized controlled trials (RCTs) that were registered with the U.S. Food and Drug Administration (FDA) with results from the published literature. They found that the findings of 23 of the 74 FDA-registered studies (31%) were unpublished. Of the 74 studies, the FDA deemed 38 (51%) "positive." All but 1 of these 38 "positive" studies were published. The FDA deemed the remaining 36 studies (49%) either "negative" (24 studies) or "questionable" (12 studies). Of these

36 studies, only 3 were published as "non-positive." The remaining 33 either were not published (22 studies) or were published as positive (11 studies). Overall, 48 of 51 published studies were reported as having yielded positive results (94%), although only 38 of the 74 FDA-registered studies actually yielded positive results (51%). Moreover, the mean effect size for published studies was 0.37, whereas the mean effect size for unpublished studies was significantly lower (0.15).

Perhaps the most disconcerting piece of evidence from this analysis was the evaluation of publication practices. First, the effect sizes in published journal articles exceeded the effect size from the FDA reviews of the same studies. The increases in effect sizes ranged from 11% to 69% with a median of 32%. Furthermore, there were instances of (a) presenting data from only one site of a multicenter trial when that site produced positive results but the other sites did not, (b) failing to perform intention-to-treat analyses on the entire original sample, as specified by the approved study protocol, and instead presenting "efficacy" data on a select subsample, and (c) failing to exclude from the analyses "outlier" sites, at which there were significant deviations from the approved study protocol if the exclusion of these sites would have nullified the overall efficacy of the trial.

Another groundbreaking meta-analysis examined the effect of baseline levels of depression on the effectiveness of the six most widely prescribed antidepressants (Kirsch et al., 2008). The investigators retrieved and analyzed data from all the trials that had been submitted by pharmaceutical companies to the FDA for regulatory approval, as well as all studies from the published literature. To evaluate the clinical meaningfulness of the results, the researchers used the commonly employed criteria of either a three-point improvement in Hamilton Rating Scale of Depression (HRSD) scores or an effect size (standardized mean difference) of $d = 0.50$. Overall, the mean improvement among samples who received the drug was 9.60 HRSD points ($d = 1.24$), whereas the mean improvement among placebo samples was 7.80 HRSD points ($d = 0.92$). This 1.80-point difference ($d = 0.32$) was below the established 3-point criterion of clinical meaningfulness.

The relative effectiveness (or lack thereof) of antidepressants did not change as a function of depression severity at baseline. On the other hand, the effectiveness of placebo decreased as the severity of depression at baseline increased. The difference between drug and placebo rose as a linear function of baseline depression severity, reaching the criterion for clinical meaningfulness only for those patients with very high baseline HRSD scores (i.e., 28 or higher, with "very severe" depression defined by the American Psychiatric Association as a score of 23 or higher). However, this difference was due to the decreased effectiveness of placebo when baseline depression was high rather than any increase in the effectiveness of the antidepressants. There was virtually no difference in the improvement scores for drug and placebo for individuals with moderate depression (but note that there was only one study of individuals with moderate depression). The researchers commented that in these trials the placebo accounted for approximately 80% of the effect of the antidepressant drug, when only 50% of the effect of analgesics can be replicated by placebo.

An independent commentator (Ioannidis, 2008) wrote: "Perhaps most people given antidepressants for depressive symptoms would just need some attention from their physician and people to talk to and take some care of them. Antidepressants may be covering largely the lost placebo of human interaction and patient–physician interaction that has become so sparse in modern society" (p. 7). In support of this general assessment, researchers have found that the apparent effectiveness of both antidepressants and placebos increases as a linear function of the number of follow-up visits with the prescribing physician during a trial. In fact, there was a 0.6–0.9 point improvement in the HRSD for each visit that took place during the trial (Posternak & Zimmerman, 2007). Ioannidis (2008) added:

> If most of the antidepressant efficacy reflects simply the placebo effect, and if most people just benefit as much as the placebo effect allows, is it unethical to kill a living myth? One might argue that if the general population is informed that antidepressants are not really effective, this might demolish the benefits that we get from the placebo effect when we administer these drugs. However, is it not unethical to lie to patients that an intervention is effective when it is not?
>
> *(Ioannidis, 2008, p. 7)*

Although the findings of the meta-analysis by Kirsch et al. (2008) are very hard to dismiss, the limitation of having included only one study of individuals with moderate depression (HRSD of 14 to 18) and no studies of individuals with mild (HDRS score of 8 to 13) or severe (HRSD score of 19 to 22) depression was important. To address this limitation, a new meta-analysis pooled together individual data from six large placebo-controlled trials that included patients with a wide range of baseline depression scores (Fournier et al., 2010). In agreement with the findings of Kirsch et al., Fournier et al. found a significant baseline severity by treatment interaction. Specifically, for patients in the mild to moderate range, the effect size was only $d = 0.11$ (with 95% confidence interval from –0.18 to 0.41) and for patients in the severe range the effect size was $d = 0.17$ (with 95% confidence interval from –0.08 to 0.43). Only for patients in the very severe group, did the effect size approach the criterion for clinical meaningfulness (i.e., $d = 0.47$, with 95% confidence interval from 0.22 to 0.71). Again supporting the results of the Kirsch et al. meta-analysis, Fournier et al. estimated that antidepressants become more effective than placebo only for patients with very severe baseline depression scores (i.e., HRSD score of 27 or higher).

Besides effectiveness, drugs must also be evaluated for their safety. It is well established that patient adherence to prescriptions for antidepressants is low and the rates of early discontinuation are high (Gartlehner et al., 2005; Goethe, Woolley, Cardoni, Woznicki, & Piez, 2007; Hunot, Horne, Leese, & Churchill, 2007). Specifically, although most prescribing physicians recommend to their patients to stay on antidepressants for at least 6 months (Bull et al., 2002), the majority of patients discontinue antidepressant therapy during the first 30 days (42.4%), whereas only 27.6% continue for more than 90 days (Olfson, Marcus, Tedeschi, & Wan, 2006). Although part of the non-adherence might be due to factors such as misunderstanding the physician's instructions or patient forgetfulness or lack of motivation, the perceived lack of effectiveness and a wide range of adverse side-effects are also responsible. Experiencing one or more "extremely" bothersome side effects has been found to be associated with more than a doubling of the risk of discontinuation (Goethe et al., 2007).

At the top of the list of the most disconcerting side effects of modern antidepressants is the increased risk of suicidal ideation and suicide attempts. Accumulating evidence of this risk led the FDA to add a "black box" warning (the strictest type of warning) on the packaging of selective serotonin reuptake inhibitor (SSRIs) antidepressants in 2004. The warning reads in part: "Antidepressants increased the risk of suicidal thinking and behavior (suicidality) in short-term studies in children and adolescents with Major Depressive Disorder (MDD) and other psychiatric disorders." In 2005, the warning was extended to include all antidepressant drugs and in 2007 the warning was extended to young adults (aged 18 to 24 years). Although the absolute number of suicide attempts and suicides is relatively small, the risk associated with the use of antidepressants is significantly elevated compared to placebo (Healy, 2003). For example, a study in the United Kingdom found that, compared with patients who had started taking an antidepressant more than 90 days before developing (nonfatal) suicidal behavior, those who had started taking an antidepressant within 1 to 9 days earlier showed a 4.07-fold higher risk of a suicidal episode, and

those who had started taking an antidepressant within 10 to 29 days earlier showed a 2.88-fold higher risk (Jick, Kaye, & Jick, 2004). Similarly, a study in Canada found that SSRI users had a 4.8-fold increased risk of suicide during the first month after receiving their prescription compared to case-matched controls (Juurlink, Mamdani, Kopp, & Redelmeier, 2006).

Second, SSRIs have been found to increase the risk of sexual dysfunction, particularly reducing sexual desire and causing problems in achieving orgasm (Balon, 2006; Kennedy & Rizvi, 2009). A meta-analysis estimated that sexual dysfunction occurs in 25.8% to 80.3% of patients on SSRIs compared to placebo (Serretti & Chiesa, 2009).

Third, SSRIs have been found to increase the risk of bleeding of the upper gastrointestinal tract (Turner, May, Arthur, & Xiong, 2007). The risk has been found to be 3.6 times higher than for non-users but can increase 9.1-fold (de Abajo & García-Rodríguez, 2008) or 12.2-fold (Oksbjerg Dalton et al., 2003) when SSRIs are taken together with aspirin or non-steroidal anti-inflammatory drugs.

Fourth, patients on SSRIs complain of sleep disturbances. In a cohort of 2,853 women over the age of 71 years, the women taking SSRI were 2.15 times more likely than non-users to sleep for 5 hours or less, 2.37 times more likely to experience sleep efficiency of 70% or less, 3.99 times more likely to experience a latency of 1 hour or more before falling asleep, and 1.75 times more likely to experience eight or more long wake episodes (Ensrud et al., 2006). Importantly, in this study sleep variables were assessed objectively by actigraph and analyses were adjusted for a multitude of potential confounders (i.e., age, race, health status, social support, physical activity, alcohol intake, medical conditions, history of depression, functional status, current benzodiazepine use, current use of non-benzodiazepine anxiolytic or hypnotic medications, depressive symptoms, symptoms of anxiety, cognitive function, and BMI).

Fifth, most classes of antidepressants have been known to cause weight gain both acutely and chronically. Weight gain is typically larger with tricyclic antidepressants and monoamine oxidase inhibitors but can also be significant with SSRIs (Deshmukh & Franco, 2003; Schwartz, Nihalani, Jindal, Virk, & Jones, 2004). Furthermore, a population study in Canada (Kisely, Cox, Campbell, Cooke, & Gardner, 2009) showed that older adults (67 years of age or older) had 34% increased risk of developing hypertension as early as 3 months after the initial prescription of SSRIs (after controlling for age, sex, socioeconomic class, schizophrenia, beta blockers, diuretics, anti-psychotics, and other drugs).

Thus, psychotherapy and pharmacotherapy, despite being recognized as the "standard" or "mainstream" approaches for the treatment of mental disorders, may not be as effective or as safe as they are commonly portrayed. Furthermore, it is important to point out that for some mental disorders, neither psychotherapy nor pharmacotherapy can presently offer any plausible preventive or treatment options. For example, in the case of Alzheimer's disease and other dementias, an expert panel convened by the National Institutes of Health in the United States (Daviglus et al., 2010) concluded that "no consistent epidemiologic evidence exists for an association with statins, antihypertensive medications, or anti-inflammatory drugs [while] data are insufficient to comment on cholinesterase inhibitors" (p. 178). Furthermore, "no known medication can be said to reliably delay the onset of Alzheimer disease" (p. 179). To the contrary, "some available evidence shows that certain medications may increase the incidence of Alzheimer disease" (p. 179; also see Plassman, Williams Burke, Holsinger, & Benjamin, 2010).

## Physical activity as an underappreciated mental health intervention

The option of physical activity interventions for promoting mental health must be evaluated in the context of the emerging concerns outlined in the previous section. In essence, the growing

concerns about the efficacy, safety, and cost of the main forms of therapy have prompted a search for other viable options. It is reasonable to propose that the conditions have now matured for the consideration and acceptance of effective, safe, inexpensive, and widely accessible alternative preventive and therapeutic modalities for a broad spectrum of mental health problems. Of the available options, physical activity seems supremely positioned due to an already-voluminous evidence base, virtual absence of adverse side effects, minimal cost, limitless global accessibility, and a wide range of collateral benefits, including those on metabolic and cardiovascular health (Phillips, Kiernan, & King, 2003).

## *Recent advances and emerging challenges*

In recent years, some important steps have been taken in bringing exercise and physical activity closer to widespread clinical application in the domain of mental health. For example, the latest (third) edition of the *Practice Guideline for the Treatment of Patients with Major Depressive Disorder*, published by the American Psychiatric Association (Gelenberg et al., 2010), states the following in regard to the role of exercise:

> For most individuals, exercise carries benefits for overall health. Data generally support at least a modest improvement in mood symptoms for patients with major depressive disorder who engage in aerobic exercise or resistance training. Regular exercise may also reduce the prevalence of depressive symptoms in the general population, with specific benefit found in older adults and individuals with co-occurring medical problems . . . If a patient with mild depression wishes to try exercise alone for several weeks as a first intervention, there is little to argue against it, provided the patient is sufficiently monitored for an abrupt worsening of mood or adverse physical effects (e.g., ischemia or musculoskeletal symptoms). The dose of exercise and adherence to an exercise regimen may be particularly important to monitor in the assessment of whether an exercise intervention is useful for major depressive disorder. If mood fails to improve after a few weeks with exercise alone, the psychiatrist should recommend medication or psychotherapy. For patients with depression of any severity and no medical contraindication to exercise, physical activity is a reasonable addition to a treatment plan for major depressive disorder. The optimal regimen is one the patient prefers and will adhere to.
>
> *(Gelenberg et al., 2010, pp. 29–30)*

Likewise, the updated edition of the *Guideline on the Treatment and Management of Depression in Adults*, issued by the National Collaborating Centre for Mental Health and the National Institute for Health and Clinical Excellence (2010) in the United Kingdom, states that "for people with persistent subthreshold depressive symptoms or mild to moderate depression," clinicians should "consider offering one or more of the following interventions, guided by the person's preference: (a) individual guided self-help based on the principles of cognitive behavioral therapy, (b) computerized cognitive behavioral therapy, or (c) a structured group physical activity program" (p. 213). Along similar lines, an evidence-based national clinical guideline for the *Non-pharmaceutical Management of Depression in Adults*, issued by the Scottish Intercollegiate Guidelines Network (2010), states that "structured exercise may be considered as a treatment option for patients with depression" (p. 10).

In other cases, although the evidence is compelling, the response from agencies with authority to issue clinical guidelines or recommendations has been tepid. A case in point is dementia and

cognitive decline. An expert panel convened to summarize the conclusions of the state-of-the-science conference on Preventing Alzheimer Disease and Cognitive Decline, organized by the National Institutes of Health (Daviglus et al., 2010) stated the following:

> Some evidence from small interventional studies and selected observational studies suggests that increased physical activity, including walking, may help maintain or improve cognitive function in normal adults. A meta-analysis of several RCTs, many with methodological limitations, concluded that data were insufficient to state that aerobic activity improves or maintains cognitive function. A small, higher-quality randomized trial of physical activity in persons with confirmed memory problems showed modest benefit in reducing cognitive decline; however, these data should be viewed as preliminary. Work is ongoing to further investigate the benefits of physical activity.
>
> *(Daviglus et al., 2010, p. 180)*

Although the language conveys very little enthusiasm, when this statement is juxtaposed to statements regarding other factors whose association to dementia and cognitive decline was also examined at the conference, one gets the impression that physical activity may be among the most promising interventions. For most other potential interventions, the panel found no indication of a beneficial effect. Nevertheless, experts have strongly criticized the conservative tone of this statement, arguing that, by considering solely large clinical trials as sources of evidence, the conclusions of the panel underestimated and misrepresented the strength of the evidence. Flicker, Liu-Ambrose, and Kramer (2011) commented that the evidence for physical activity was "discounted too cursorily for what is now a large and relatively consistent pool of animal and human data" and that "it is difficult to understand why the NIH consensus statement has taken such a cautious approach" (p. 466). Ahlskog, Geda, Graff-Radford, and Petersen (2011) characterized the approach taken by the panel as "nihilistic" (p. 877). Arguably, however, the portrayal of the evidence for physical activity was not only "cautious" and "nihilistic." In a rather remarkable excerpt, the panel clustered together physical activity and "other leisure activities," as if implying that the respective evidence pertaining to these types of "activities" were comparable in size, strength, quality, or consistency: "Preliminary evidence suggests beneficial associations of physical activity and other leisure activities (such as club membership, religious services, painting, or gardening) with preservation of cognitive function" (p. 178).

Over the years, authors from various backgrounds have expressed puzzlement, even frustration, that physical activity, despite its apparent strengths, is not being promoted more actively as a preventive or therapeutic intervention in the domain of public health (Berk, 2007; Callaghan, 2004; Daley, 2002; Donaghy, 2007; Faulkner & Biddle, 2001; Otto et al., 2007; Taylor & Faulkner, 2008; Tkachuk & Martin, 1999). The reality is that, with growing recognition over the years, new challenges have emerged. When physical activity is recommended on a national scale as an effective and safe alternative to established therapies for mild or moderate depression or when physical activity appears to show more promise than any other form of intervention for slowing the progression of cognitive decline, it is natural to stimulate broader interest and, with that, elicit some unintended consequences. These include reactions fueled by dualistic incredulity, opposition stemming from disciplinary territorialism, and, importantly, the wrath of significant financial interests. In combination, these responses raise the risk of bias tingeing the assessments of the evidence that appear in the published literature (MacCoun, 1998). As a result, the literature has been inundated with conflicting, confusing, and misleading conclusions and has become immensely challenging to evaluate, particularly for non-specialists.

## The crucial role of paradigms

It would be shortsighted, even outright erroneous, to read the following sections without placing them within a broader epistemological framework. Neither science nor science-driven clinical practice work simply on the basis of what the evidence shows. They work on the basis of paradigms, as defined by Kuhn (1962/1996). Paradigms do not coexist peacefully and do not transition smoothly from one to the next. They battle each other. According to Kuhn, a dominant paradigm imposes "immense restriction of the scientist's vision," to the point that science becomes "increasingly rigid" (p. 64). Consequently, there is always "considerable resistance to paradigm change" (p. 64). Even when scientists "begin to lose faith [and] consider alternatives," they "do not renounce the paradigm that has led them into crisis" (p. 77). The main reason is that "as in manufacture so in science-retooling is an extravagance to be reserved for the occasion that demands it" (p. 76).

Kuhn (1962/1996) emphasized that the transition from an old paradigm to a new one "is far from a cumulative process" (p. 84). Instead, a crisis that forces a field into a search for alternatives essentially induces "a reconstruction of the field from new fundamentals, a reconstruction that changes some of the field's most elementary theoretical generalizations as well as many of its paradigm methods and applications" (p. 85). Because of the radical nature of transitions, such "reconstructions" progress slowly and usually amid considerable tension. Kuhn further suggested that, because the paradigm is such an integral component of the careers of established scientists, researchers who attempt the "reconstruction of the field from new fundamentals" are "almost always . . . either very young or very new to the field whose paradigm they change" (p. 90). These characteristics give them an advantage because they entail that these scientists are "little committed by prior practice to the traditional rules of normal science" and, as a result, they "are particularly likely to see that those rules no longer define a playable game and to conceive another set that can replace them" (p. 90). When a scientific revolution to replace an existing paradigm begins, it is "often restricted to a narrow subdivision of the scientific community" (p. 92).

A relevant example that illustrates the validity of these points is the recognition of the role of exercise in cardiovascular health. Writing in a major medical journal as recently as 1979, Kannel and Sorlie asserted that "it does not seem likely that exercise programs can make as great an impact on incidence of cardiovascular disease as can control of blood pressure, the cigarette habit, obesity, or hyperlipemia" (p. 860). Today, physical activity is widely regarded as one of, if not the, most powerful interventions for the promotion of cardiovascular health and physical activity recommendations are published jointly by the American Heart Association and the American College of Sports Medicine (Haskell et al., 2007). It is therefore doubtful that a similar statement would find its way into a major medical journal.

## Simpler times

When the first reviews on the association between physical activity and mental health appeared in the literature, the evidence base consisted only of small-scale correlational and a few quasi-experimental studies. Nevertheless, there was hopeful enthusiasm on the part of exercise scientists that, eventually, the evidence would demonstrate that exercise and physical activity have important roles to play in this important domain. With the thin red line separating scientific impartiality from advocacy being somewhat fuzzy, disagreements were bound to emerge. For example, an exercise scientist asserted that "the 'feeling better' sensation that accompanies regular physical activity is so obvious that it is one of the few universally accepted benefits of exercise" (Morgan, 1981, p. 306). At the same time, based on the same limited data, a psychiatrist and a stress physiologist concluded that, although "the psychological benefits of fitness training,

especially jogging, have been propagandized by the popular press" (Folkins & Sime, 1981, p. 373), the "status of theorizing about the processes that might explain physical fitness training effects can best be described as a potpourri of speculations" (p. 374) and the "studies of physical fitness effects on psychological health are poorly designed" (p. 386). A little later, another psychiatrist agreed, stating that "the enthusiastic support of exercise to improve mental health has a limited empirical basis and lacks a well tested rationale" (Hughes, 1984, p. 76). These early disagreements can be easily attributed to the fact that one side was basing its statements on belief or intuition whereas the other on a more dispassionate assessment of the scant evidence.

## *Exercise as analogous to complementary healers*

Despite substantial growth in the evidence base during the 1980s and 1990s, views on the role of physical activity in mental health continued to diverge. In 1996, a landmark report of the Surgeon General of the United States on the relationship between physical activity and health included the following statement:

> The literature suggests that physical activity helps improve the mental health of both clinical and nonclinical populations. Physical activity interventions have benefitted persons from the general population who report mood disturbance, including symptoms of anxiety and depression, as well as patients who have been diagnosed with nonbipolar, nonpsychotic depression. These findings are supported by a limited number of intervention studies conducted in community and laboratory settings . . . The psychological benefits of regular physical activity for persons who have relatively good physical and mental health are less clear.
>
> *(United States Department of Health and Human Services, 1996, p. 136)*

In 1999, the 458-page report of the Surgeon General on mental health did not mention physical activity among the recognized methods of treatment for anxiety and depression, which only included psychotherapy and pharmacotherapy. Physical activity was mentioned as one of an "ever-expanding list" of "informal" interventions for coping with stressful life events, alongside "religious and spiritual endeavors" and "complementary healers" (United States Department of Health and Human Services, 1999, p. 232). Citing the earlier report, some possible benefits were acknowledged. However, this was immediately followed by a caveat about the poor methodological quality of the evidence base:

> Physical activities are a means to enhance somatic health as well as to deal with stress. A recent Surgeon General's Report on Physical Activity and Health evaluated the evidence for physical activities serving to enhance mental health. Aerobic physical activities, such as brisk walking and running, were found to improve mental health for people who report symptoms of anxiety and depression and for those who are diagnosed with some forms of depression. The mental health benefits of physical activity for individuals in relatively good physical and mental health were not as evident, but the studies did not have sufficient rigor from which to draw unequivocal conclusions.
>
> *(United States Department of Health and Human Services, 1999, p. 232)*

## *"There is no evidence"*

To date, there is only one empirical study that examined the level of awareness among professionals in clinical psychology of the research literature on physical activity and mental health.

Faulkner and Biddle (2001) conducted interviews with 21 directors of doctoral programs in clinical psychology in the United Kingdom. The interviews showed that physical activity is, for the most part, ignored in clinical psychology curricula. The authors attributed this phenomenon to a certain degree of disciplinary bias, fueled by "a continued adherence to a dualistic notion of mental illness and mental health" (Faulkner & Biddle, 2001, p. 442). What was most relevant to the present review was the finding that the program directors were unaware of research evidence supporting a role of physical activity in the domain of mental health. Faulkner and Biddle (2001) wrote: "Awareness of the exercise and mental health literature was extremely limited, with most participants being unfamiliar with existing research" (p. 439). This conclusion was based on statements by program directors, such as the following:

> There is no evidence. Although you might find the odd paper which says that exercise is effective here and there, as far as treating clinical problems, populations with psychological and psychiatric problems, to my knowledge there is no evidence.
>
> *(Faulkner & Biddle, 2001, p. 439)*

Another program director wondered: "We might want to ask the question, if there's evidence for exercise, why is noone mentioning it?" (p. 439). Other respondents questioned the quality of the extant evidence but were unable to refer to specific problems because this concern was based on suspicion rather than actual first-hand knowledge: "the participants were unable to offer critical insight into the nature of existing research because they were unaware of its existence in the first place" (pp. 439–440).

## *Perhaps "a bit harsh" but at least a stimulus for improvement*

Conflicting assessments continued to appear in the new millennium. A much-discussed issue was depression. Following a systematic review of the evidence on the effects of exercise on depression, commissioned by the Somerset Health Authority in the United Kingdom, a panel of exercise scientists concluded that "overall, the evidence is strong enough for us to conclude that there is support for a causal link between physical activity and reduced clinically defined depression. This is the first time such a statement has been made" (Biddle, Fox, Boutcher, & Faulkner, 2000, p. 155). At the same time, an epidemiologist and a psychiatrist, based on a meta-analysis on the same topic, despite finding that exercise had a substantial positive effect (mean effect size of 1.1) and was no less effective than (much costlier) cognitive therapy, concluded that "the effectiveness of exercise in reducing symptoms of depression cannot be determined because of a lack of good quality research on clinical populations with adequate follow up" (Lawlor & Hopker, 2001, p. 1).

One possible reason for these conflicting evaluations is the scope of the evidence that was considered in each case. Biddle et al. (2000) examined a broad range of sources, spanning the gamut from epidemiologic studies to experiments and neurobiological data. On the other hand, Lawlor and Hopker (2001) considered only RCTs. Furthermore, for Lawlor and Hopker, the evaluation of methodological rigor was based solely on three criteria, namely, "whether allocation was concealed and intention to treat analysis was undertaken, and whether there was blinding" (p. 2). On the basis of these criteria, they judged most studies as being "of poor quality" (p. 3) or "low quality" (p. 5). Closer examination of their criteria, however, reveals a rather peculiar combination of punctiliousness and superficiality. For example, concealment of treatment allocation was judged as "adequate" only if a report specifically mentioned that there was "central randomization at a site remote from the study; computerized allocation in which records are in a locked, unreadable file that can be accessed only after entering patient details; the drawing of

sealed and opaque sequentially numbered envelopes" (p. 2). Some of these criteria have been characterized as "arbitrary" (Blumenthal & Ong, 2009, p. 97). At the same time, more consequential threats to internal validity received no attention (Ekkekakis, 2008; Ekkekakis & Backhouse, 2009).

Given the fact that the Lawlor and Hopker (2001) meta-analysis covered studies published until 1999 and the first Consolidated Standards of Reporting Trials (CONSORT) were published in 1996 (Begg et al., 1996), it is unsurprising that methodological details, such as the exact method of randomization and concealment of treatment allocation, were not reported with the greatest degree of specificity. Lawlor and Hopker found that "in no study was treatment allocation described, and contact with authors established that allocation might have been adequately concealed in only three [of 14] studies" (p. 3). In other words, the concealment was judged as inadequate for 79% of the studies. For comparison, prior to the publication of the original CONSORT statement, allocation concealment was judged as "unclear" for 79% of RCTs published in the same highly regarded medical journal in which Lawlor and Hopker published their analysis (Moher, Jones, & Lepage, 2001). Therefore, although highlighting departures from modern reporting guidelines may give a semblance of scientific stringency and provide a basis for summarily dismissing studies as being of low quality, it also overlooks the fact that the standards of this literature were not much different from those of the broader medical literature of the pre-CONSORT era.

The conclusions and critical assessment of the literature by Lawlor and Hopker (2001) prompted a range of responses from researchers working in this field. Defending their approach, Lawlor and Hopker cited an earlier analysis by Schulz, Chalmers, Hayes, and Altman (1995), which found that "trials that reported either inadequate or unclear concealment methods yielded estimates of [odds ratios] that were exaggerated by an average of 41% or 30%, respectively, compared with estimates of [odds ratios] derived from trials that apparently had taken adequate steps to conceal treatment allocation" (p. 411). However, as Callaghan (2004) pointed out, even under the worst-case scenario of inadequate concealment, the exercise effect may be attenuated but remains quite substantial: "on closer scrutiny it would be more accurate to conclude, in my view, that the evidence is weakened but not invalidated. In their review Lawlor and Hopker report an average effect size for exercise . . . of 1.1. Taking the upper end of Schulz et al.'s figure, we can reduce this effect size by 40% to account for methodological weaknesses. This leaves an effect size of 0.66, medium by widely cited estimates" (p. 478).

Other authors criticized the narrow view of the evidence but also conceded that the quality of the evidence was still lacking compared to the strict standards typically expected of RCTs that form the foundation for evidence-based clinical practice:

> It would appear that this assessment is a bit harsh, insofar as there are many studies that, taken together, offer considerable evidence for the benefits of exercise in reducing depression in clinical populations. However, it also is true that there are limited data from well designed clinical trials, which are often considered the gold standard for evaluating the effectiveness of a new therapy . . . Because of the limited data currently available, it could be stated that if the prescription of exercise for [major depressive disorder] required approval from the Food and Drug Administration, it probably would not pass current standards.
>
> *(Brosse, Sheets, Lett, & Blumenthal, 2002, p. 754)*

Despite its arguably inordinate emphasis on technicalities that may be of little practical consequence while missing other crucial problems, the review by Lawlor and Hopker (2001) can be

credited with a sensitization of researchers to the need to adhere closely to reporting guidelines and describe methodological details with greater clarity. For example, using language strongly reminiscent of the Lawlor and Hopker criteria for adequate allocation concealment, Dunn, Trivedi, Kampert, Clark, and Chambliss (2005) wrote that "randomization was implemented with sequentially numbered, opaque, sealed envelopes" (p. 2). Similarly, Blumenthal et al. (2007) wrote that "randomization was performed centrally by computer" and "patients were provided with sealed envelopes containing their group assignment" (Blumenthal et al., 2007, p. 588). As a result, subsequent updates of the meta-analysis have identified more studies that satisfied the quality assessment criteria. The total number, however, remains low. Specifically, out of 25 studies included in the analysis in 2009 (Mead et al., 2009) and 30 that were included in 2012 (Rimer et al., 2012), only three and four, respectively, satisfied all three criteria (i.e., adequate allocation concealment, intention-to-treat analyses, and blinded outcome assessors). Limiting the analysis to such a small group of studies rendered the average effect not statistically significant. Furthermore, although the analysis indicated that "when compared with other established treatments (cognitive behavioral therapy and antidepressants), there was no difference between exercise and the established intervention," the authors insist that "outstanding uncertainties remain about how effective exercise is for depression, mainly because of methodological considerations" (Rimer et al., 2012, p. 17).

## *Distinguishing findings from their interpretation*

Placing emphasis on methodological stringency generally tends to give reviews the semblance of a "strict but fair" impartiality, which is inherently appealing to scientists. However, closely parsing the language of critical reviews can be very revealing. For example, the conclusion of a recent meta-analysis on the effects of exercise on depression among individuals who satisfied diagnostic criteria for depression was:

> The results of this systematic review and meta-analysis suggest that exercise at most has a small benefit in relieving symptoms of depression in patients with clinically diagnosed depression in the short term, based on the [standardized mean difference] of –0.4 which is within the range considered to represent a small effect (0.2 to 0.5).
>
> *(Krogh, Nordentoft, Sterne, & Lawlor, 2011, p. 535)*

First, Cohen (1992) did not define a "small" effect as a range from 0.2 to 0.5 standard deviation but rather as a point estimate of 0.5 standard deviation: "the small, medium, and large [effect sizes] are $d = .20$, .50, and .80. Thus, an operationally defined medium difference between means is half a standard deviation" (Cohen, 1992, p. 157). An effect size of 0.4 is closer to 0.5 (and, therefore a "moderate" effect) than it is to 0.2 (or "small").

Second, and more importantly, the non-exercise comparison groups included active treatments, such as pharmacotherapy with regular doses of sertraline hydrochloride, group psychotherapy, and non-aerobic forms of exercise, all of which have been shown effective for reducing depression. By using these comparison groups, studies that showed substantial exercise-induced decreases in symptoms of depression were entered in the analysis as having a null or even a slightly negative effect, lowering the overall pooled standardized mean difference.

Third, the reviewers noted that they found "no evidence that this small effect lasted beyond the duration of the exercise program" (Krogh et al., 2011, pp. 535–536). However, as it has justly been noted, "there is no reason to believe that exercise training would continue to benefit patients after they discontinued exercise any more than patients who stopped taking their

antidepressant medication would continue to benefit from medication if they discontinued their medication" (Blumenthal & Ong, 2009, p. 98).

Fourth, the reviewers committed a classic logical fallacy: George has a beard. Goats have beards. Therefore, George is a goat. Now, substitute "George" for "exercise" and "goat" for "placebo effect": "The reduced effect in trials of longer duration might suggest that any effect of exercise is largely placebo in nature, since placebo effects tend to diminish with time" (Krogh et al., 2011, p. 536). There are, of course, numerous other, arguably more likely, scenarios for an effect whose magnitude appears to be reduced over time, before one must arrive at the conclusion that the culprit is the placebo effect.

## *RCTs as the sole credible source of scientific evidence*

Starting with the oft-cited Lawlor and Hopker (2001) review, it has become common practice in systematic reviews dealing with physical activity and various aspects of mental health to limit study selection only to randomized clinical trials, then reject most of these trials on the basis of various methodological weaknesses, then find a substantial average effect size from the remaining trials, but finally dismiss this finding because these remaining trials are characterized as too few or too small. On the surface, focusing exclusively on RCTs appears reasonable; after all, this is the standard method used in evaluating the effectiveness of drugs.

On the other hand, this approach essentially sets up an unwinnable game. There are no hard rules for either the number or the size of RCTs that can be thought to constitute an adequate evidentiary basis. So, a reviewer could simply declare, for example, that five RCTs, despite showing strong and consistent positive results, are too few or that their sample sizes are not above a certain arbitrary number that seems large (e.g., 100 participants), regardless of whether the trial was adequately powered. Likewise, a reviewer can always call for even larger and even longer RCTs. Although this may seem hard to refute, there are limits to what can be funded. It should be kept in mind that, unlike trials funded by the pharmaceutical industry, trials of exercise and physical activity are funded mostly by governmental sources and have considerably smaller budgets. Moreover, as the evidence for the benefits of physical activity mounts, governmental funding agencies become increasingly reluctant to approve trials that include the randomization of vulnerable individuals, such as mental health patients or elderly individuals with dementia or cognitive impairment, to control groups (Flicker et al., 2011).

At this point, readers of the literature are confronted with a very confusing situation. Expert panels examine the same evidence base and arrive at conflicting conclusions. For patients and clinicians looking for guidance, it is very difficult to understand how this is possible or what they should do. One apparent cause for the disagreements is that different reviewers apply different study inclusion and exclusion criteria. Some limit their scope to RCTs because this type of design has the potential to provide the most reliable evidence. Others also consider observational and quasi-experimental studies, as well as basic science research, thus drawing information from a much broader and more diverse range of sources. Other possible sources of disagreement are the criteria for evaluating study quality, as well as how consequential each methodological weakness is interpreted to be. Even after taking these different approaches to systematic reviewing into account, however, disagreements are still likely to occur because, ultimately, the process of evaluating the evidence is impossible to disentangle from the element of subjectivity.

As one example, an expert panel, convened under the direction of the Centers for Disease Control and Prevention Healthy Aging Research Network, conducted an extensive review of the evidence on whether exercise and physical activity interventions can maintain or improve the cognitive function of older community-dwelling adults (Snowden et al., 2011). The

conclusion of the panel was that "no intervention-outcome pairings had good quality overall" (p. 714) and, therefore, there is "insufficient evidence of cognitive benefits from exercise or physical activity interventions in older adults" (pp. 714–715). As in every other review focusing solely on RCTs, the panel recommended "larger-scale, rigorous clinical trials with longer follow-up periods" (p. 715).

At the same time, another panel of experts (from psychiatry and neurology), after examining evidence that also included neurobiological mechanisms of cognition, concluded that "aerobic exercise is associated with a reduced risk of cognitive impairment and dementia" and "may slow dementing illness." This is because of "two plausible biologic pathways" including attenuating the progression of neurodegenerative processes and reducing vascular risk factors that contribute to dementia risk, particularly via small vessel disease. For these reasons, the authors recommended that "moderate-intensity physical exercise should be considered as a prescription for lowering cognitive risks and slowing cognitive decline across the age spectrum" (Ahlskog et al., 2011, p. 882).

At this point, it should be clear that the process of systematic reviewing is not working as intended. If it did, there would not be as many systematic reviews whose conclusions stand in stark contrast to each other. Therefore, a reasonable proposal is to encourage the standardization of the process and, importantly, increase the transparency by which conclusions are reached and recommendations are made. The practice of summarily rejecting a body of research should be replaced by a specific account of the perceived methodological quality of each study that was considered, the reasons for this judgment, and the weight allocated to each study in reaching a final conclusion. The field of evidence-based medicine has made considerable strides in developing quasi-standardized guidelines, such as the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) protocol. The GRADE approach acknowledges that "the quality of evidence represents a continuum" (Balshem et al., 2011, p. 404) and consistent evidence from observational studies has its rightful place. Furthermore, it appears that, at a minimum, adhering to the GRADE process can reduce some of the arbitrariness and enhance the transparency of systematic reviews in the field of physical activity and mental health:

> There will be cases in which competent reviewers will have honest and legitimate disagreement about the interpretation of evidence. In such cases, the merit of GRADE is that it provides a framework that guides one through the critical components of this assessment and an approach to analysis and communication that encourages transparency and an explicit accounting of the judgments involved.
>
> *(Balshem et al., 2011, p. 405)*

## *There is pro-exercise bias too, particularly in the exercise literature*

Advocacy is seldom a good guide in science and it is true that, over the years, many statements about the mental health benefits of physical activity have tended to "anticipate rather than reflect the accumulation of strong evidence" (Salmon, 2001, p. 36). Therefore, readers would be well served to apply the same degree of scrutiny and skepticism to assessments of the evidence that seem to favor a role for physical activity in mental health as to those that question it.

As an example, one meta-analysis concluded that its results "provide Level 1, Grade A evidence supporting the use of exercise for the alleviation of depressive symptoms" (Rethorst, Wipfli, & Landers, 2009, p. 506) and another that its results represent "Level 1, Grade A evidence for using exercise as a treatment for anxiety disorders" (Wipfli, Rethorst, & Landers, 2008, p. 401). Both meta-analyses were published in highly regarded exercise science journals.

The "Level 1, Grade A" (i.e., top-level) classification refers to a now-obsolete grading system proposed by Guyatt, Cook, Sackett, Eckman, and Pauker (1998) for the evaluation of evidence for different types of antithrombotic medications. According to the logic of this grading system, to achieve the highest level of recommendation, evidence must come from "meta-analysis or a large randomized trial with consistent results" (p. 444S). Thus, for "Grade A," the evidence base must exhibit the following characteristics: "methods strong, results consistent, RCTs, no heterogeneity" (Guyatt et al., 1998, p. 442S, see Table 1). The authors further explained that "one approach to arriving at a best estimate is to conduct rigorous systematic reviews and meta-analyses of RCTs. Another is to employ the results of a single large study if it provides a more accurate estimate of the treatment effect" (p. 442S). Under "Grade A" evidence, there are two levels, 1 and 2. What distinguishes the two levels is whether the effects are clearly beneficial ("Level 1") or equivocal ("Level 2," i.e., it is "uncertain whether the benefits outweigh the risks").

What distinguishes "Grade A" from "Grade B" evidence is the presence (for "Grade A") or absence (for "Grade B") of consistency in the results of the different studies. Thus, for "Grade B" evidence, the criteria are: "methods strong, results inconsistent, RCTs, heterogeneity" (p. 442S, see Table 1). Within "Grade B" evidence, there are again two levels, with "Level 1" evidence requiring "effect clear" (i.e., "that benefits outweigh risks"), and "Level 2" evidence established when the "effect [is] equivocal" (i.e., "uncertainty whether benefits outweigh risks").

In interpreting Guyatt et al.'s (1998) grading system, Wipfli et al. (2008) wrote that, because "large-scale randomized trials are often extremely costly and time consuming to conduct," meta-analysis offers a way to "get around" this problem by "combining results of Level 2 studies, which are smaller randomized, controlled trials" (p. 394). Along the same lines, Rethorst et al. (2009) argued that "Level 1, Grade A evidence can be provided either through one large randomized controlled trial or through a meta-analysis of smaller (Level 2, Grade B) randomized controlled trials" (p. 492). In turn, "Level 2, Grade B studies are similar to Level 1, Grade A studies in that they are randomized controlled trials. However, these studies have smaller sample sizes and thus are susceptible to type II errors" (pp. 492–493).

It seems clear that this interpretation is inconsistent with the logic of Guyatt et al.'s (1998) grading system. For Guyatt et al., the key question was that of consistency. The best case scenario would be to have multiple RCTs yielding consistent results: "Investigators will generate the strongest recommendations when RCTs yielding consistent results (grade A evidence) are available" (p. 442S). If there is only a single RCT, it may still be considered as yielding Grade A evidence ("we classify recommendations as grade A if they are based on even a single RCT"), as long as the RCT is "large," so that its estimate of treatment effect is "accurate." If the single RCT is small, the confidence interval around the point estimate of the treatment effect would be large and, if the lower bound of the confidence interval approaches zero, then there would be more "uncertainty" about the benefits of treatment. Likewise, if multiple RCTs have yielded inconsistent results, especially if some approximate a null effect, there would again be "uncertainty" about the benefit:

> When several RCTs yield widely differing estimates of treatment effect for which there is no explanation (we label this situation "heterogeneity present"), the strength of recommendations from even rigorous RCTs is weaker (grade B evidence) . . . We must acknowledge the heterogeneity of results across studies. In doing so, any recommendation would move from grade A to grade B.
>
> *(Guyatt et al., 1998, p. 442S)*

Furthermore, a crucial point was that for evidence to be classified as either "Grade A" or even "Grade B," the methods must be "strong." It is clear that the random allocation of participants to conditions is not a sufficient condition for a methodology to be labeled "strong." Even if using a randomized design, a methodology would not be characterized as "strong" if treatment allocation was not adequately concealed, assessors were not blinded (particularly when the outcomes are subjective and their assessment is susceptible to bias), when there were large losses to follow-up, when there was no intention-to-treat analysis, when the trial stopped earlier than scheduled, or when there was publication bias (e.g., failure to publish non-significant results).

With this in mind, it is important to point out that one of the aforementioned meta-analyses did not include ratings of quality (Wipfli et al., 2008), whereas the other identified only 5 studies (of 58) with allocation concealment and only 3 studies with intention-to-treat analyses (Rethorst et al., 2009). Moreover, as an indirect indicator of study quality, the authors reported that only 24% (12 of 49) of the studies in one analysis provided adequate information to calculate an exercise dose and approximately half of the effect sizes were derived from studies that did not report the intensity of exercise used (Wipfli et al., p. 403). In the other analyses, only 21% (12 of 58) of the studies provided adequate information to calculate an exercise dose and fewer than half of the studies (24 of 58) reported adequate information about the exercise intensity used (Rethorst et al., 2009). Therefore, it seems that it would not be appropriate to assume that the methods were "strong," which is a requirement for both "Grade A" and "Grade B" evidence.

## *Research in the era of sloppy journalism and social media*

With access to the vast majority of scientific journals still being subscription-based, the privilege of inspecting the "fine print" of published studies is restricted only to those inhabiting the ivory towers of academia. The rest of society is informed via the media. This process is usually mediated by the press offices of universities and research centers. Their role is a crucial one because, unlike scientific journals, which are read by only a few hundred specialists, media pieces are disseminated to millions of people worldwide, including patients in need of treatment options. With more patients becoming active participants in health care decisions, the information reaching them through the media can sway them toward asking their health care providers for one option versus another. It is, therefore, extremely important to provide patients with information that accurately depicts the latest research findings.

Conceivably, however, if the media were provided with inaccurate or misleading information by an ostensibly credible source such as a university press office, it seems fairly clear that most media organizations have neither the technical ability nor the time and inclination to do their own independent fact checking. Therefore, the misinformation would spread at an incredible speed throughout the globe, creating a false impression that would be impossible to contain or ever undo.

As any marketing specialist would attest, "contradictory counter-messaging" can be an extremely powerful tool. For many years, the tobacco industry deliberately used this technique to its advantage (Landman, Cortese, & Glantz, 2008; Smith, 2006) by disseminating a seemingly valid, supposedly science-based, counter-message to the ongoing anti-smoking social marketing campaigns targeting the addictive properties of nicotine. For a smoker who is unwilling or unable to conduct an independent analysis and evaluation of the scientific data, one "science" appears no different from the other "science" and, therefore, the message that nicotine is addictive seems no more valid than the message that it is not. In fact, the "contradictory counter-message" (i.e., that nicotine is non-addictive) does not even have to be believed. Even its mere presence can accomplish the objective, which is to create confusion: many in the public are led to believe

that, since there is apparent inconsistency in scientific findings, the addictive properties of nicotine are doubtful. In the end, it really does not matter what the actual research has shown. The battle is fought in the arena of public opinion, following very different rules than the rules of science. For these reasons, the portrayal of research findings in the media can create an additional challenge, a new layer of complexity that readers seeking reliable information have to confront.

On June 6, 2012, a university press office released a statement describing a study that had just been completed. From the opening paragraph, the statement made it clear that the study had been motivated by the clinical practice guidelines for the treatment of depression that have been in place in the United Kingdom (see the earlier section entitled "Recent advances and emerging challenges"): "Current clinical guidance recommends physical activity to alleviate the symptoms of depression. However, new research published today in the [*British Medical Journal*], suggests that adding a physical activity intervention to usual care did not reduce symptoms of depression more than usual care alone." Of course, the phrase "physical activity intervention" is understood by most people to mean that, in one experimental condition, people did physical activity, whereas in the other (the control) they did not. Nowhere in the press release is there any specific information about what exactly the "physical activity intervention" consisted of and nowhere is it described as anything other than a "physical activity intervention."

The lead researcher was quoted as saying "Our intervention was not an effective strategy for reducing symptoms. However, it is important to note that increased physical activity is beneficial for people with other medical conditions such as obesity, diabetes and cardiovascular disease." The antithesis created by the use of "however" clearly implies that, unlike the benefits of physical activity for obesity, diabetes, and cardiovascular disease, physical activity is ineffective for reducing symptoms of depression. Another senior researcher further reinforced this interpretation by saying, in no uncertain terms, that "exercise and activity appeared to offer promise as one such treatment, but this carefully designed research study has shown that exercise does not appear to be effective in treating depression."

As was entirely predictable, the British press highlighted the story, as it always happens when a new study contradicts what was believed to be established knowledge up to that point. In an interview to the British Broadcast Corporation (BBC), a senior researcher said, "exercise is very good for you, but it's not good for treating people with what was actually quite severe depression." Furthermore, "that buzz we all get from moderate intensity of exercise is certainly acknowledged but it's not sustained and it's not appropriate for treating people with depression."

A professor of public health not associated with the study told the BBC that "this is a huge disappointment because we were hoping exercise would help lift depression." On the same day, *The Guardian*, a major British newspaper carried the story under the title "Exercise doesn't help depression, study concludes." The opening sentence read: "A study into whether physical activity alleviates the symptoms of depression has found there is no benefit." In another newspaper, *The Telegraph*, under the title "Exercise fails to lift clinical depression," the medical correspondent explained that "exercise should not be 'prescribed' to people with clinical depression, according to a study which found it did nothing to improve their moods."

Only the BBC article hinted at the story behind the story, namely the implications of the study and the associated press release for the role of physical activity in public health policy in the United Kingdom. Reminding readers that the study had been funded by government (i.e., taxpayer) money, the BBC predicted that the findings are "likely to be taken into account when [the National Institute for Health and Clinical Excellence] next reviews its guidelines." A population that paid a substantial sum of money for the study to be carried out and is now convinced by the press release and subsequent media buzz of the ineffectiveness of physical activity is unlikely to accept the continuation of the use of physical activity as a treatment for depression

through the national health care system. The National Health Service (NHS) responded a day later, cautioning readers that "despite what several headlines have suggested, new research has not re-examined the effect of exercise on depression" and that "the results do not support the view that exercise is 'useless' for treating depression, as some news sources have suggested." However, it was clearly too little, too late. By that time, the headlines had traveled around the world and reverberated through the social media.

As noted earlier, in a scenario like this, what the actual study showed is ultimately rendered irrelevant. In public opinion, perception is reality and the perception has clearly been formed. Returning to the study to dissect methods and results is likely to strike most people as too technical, too difficult, too confusing, and, after a while, essentially pointless. Depressed patients visiting their primary care physicians are unlikely to accept any advice for physical activity since all major media organizations declared it ineffective. They are also unlikely to have much patience for explanations on what the study did or did not show.

Anyone with the patience to juxtapose the press release and media reports with the actual research publication would be surprised by the extent to which the evidence was distorted. As the NHS pointed out, the study did not investigate the effects of exercise or physical activity on depression. More accurately, although the originally stated aim of the study was, in fact, to "evaluate, in general terms, whether physical activity can be an effective treatment for depression within primary care" (Baxter et al., 2010, p. 6), a series of significant failures in the conceptualization and design of the study precluded the accomplishment of this goal.

The study used a large sample ($N = 361$) of patients who could be characterized as having at least "mild" depression on the basis of scores on the Beck Depression Inventory (i.e., at least a score of 14, on a scale where 0–13 represents minimal depressive symptoms, 14–19 mild, 20–28 moderate, and 29–63 severe depressive symptoms).

Beyond the large sample size, however, most aspects of the study were severely flawed. The patients were randomized to one of two conditions. In one condition, patients received usual care, which within the NHS means the full range of available treatments, including access to antidepressant medications, counseling, referral to exercise on prescription schemes, and secondary care mental health services. In the other condition, in addition to usual care, the patients received an intervention designed to promote physical activity. It should be clear, therefore, that, by design, the study incorporated multiple confounding factors, whose interactions with physical activity were unknown and left uncontrolled. Although recruitment was stratified by the baseline use of antidepressants and an extremely crude measure of physical activity (i.e., self-reported number of days per week of at least 30 minutes of "moderate" or "vigorous" activity), the researchers did not account for the use of other resources, such as counseling.

Furthermore, the researchers admitted initially expecting that only "10% of the sample would have been taking antidepressants at baseline and had intended to exclude them from the main comparative analysis" (p. 4). However, over 50% of the sample was already on antidepressants at baseline, so they "decided to include everyone in the analysis" (p. 4) and, instead, control for antidepressant use statistically, as a covariate. However, the simultaneous exposure of the majority of the intervention group to physical activity and antidepressants represents a confound, and a fatal design flaw, which cannot be remedied by resorting to a statistical control of antidepressant use. First, it is presumed that at least part of the antidepressant effect of physical activity is due to changes in the brain's serotonergic network, the same network targeted by the antidepressant medication. The effects of the interaction of these two parallel interventions on this network remain unclear. Second, it has been found that, when these two treatments are administered in tandem, the antidepressants may undermine the effects of physical activity (Babyak et al., 2000). One possibility is that this interference may be mediated by cognitive processes:

> One of the positive psychological benefits of systematic exercise is the development of a sense of personal mastery and positive self-regard, which we believe is likely to play some role in the depression-reducing effects of exercise. It is conceivable that the concurrent use of medication may undermine this benefit by prioritizing an alternative, less self-confirming attribution for one's improved condition. Instead of incorporating the belief "I was dedicated and worked hard with the exercise program; it wasn't easy, but I beat this depression," patients might incorporate the belief that "I took an antidepressant and got better."
>
> *(Babyak et al., 2000, p. 636)*

The researchers indirectly acknowledged that the statistical control for antidepressant use was only a necessary but imperfect compromise for dealing with the confound, by disclosing that they had originally intended to exclude all participants already on antidepressants. Of course, they should have also done the same with all participants receiving counseling and all those who had opted to use secondary care mental health services. These parallel treatments could also directly confound the effects of physical activity.

The intervention that was used to increase physical activity was also problematic. First, it should be pointed out that the guidelines of the National Institute for Health and Clinical Excellence (2010) specified "a structured group physical activity program" supervised by a physical activity facilitator with a Bachelor's or a Master's degree (the rationale for this was detailed on p. 212 of the guideline document). Instead, the intervention was designed to increase general (lifestyle), unsupervised, physical activity, which is a very different type of intervention.

The intervention to increase physical activity was based on a theory that happens to be in vogue in exercise psychology, namely, self-determination theory. Since no specific explanation was offered for why this particular theory was deemed most appropriate for this particular intervention or this particular population, other than that it "seemed particularly relevant" (Haase, Taylor, Fox, Thorp, & Lewis, 2010, p. 86), it is probably safe to assume that a different theory would have been used had the study been designed 5 or 10 years ago, when different theories were in vogue (e.g., the transtheoretical model or social-cognitive theory).

Importantly, this intervention had not been previously tested for its effectiveness or appropriateness in this population. Thus, it is entirely plausible that some of the precepts of self-determination theory might lead to counterproductive effects. For example, it has been found that some participants, due to lack of knowledge and experience with physical activity, feel insecure about making decisions (e.g., selecting an appropriate intensity) and would prefer to relinquish control to a trained professional. Being left with no choice but to decide for themselves, these individuals feel less confident and report reduced pleasure during physical activity (Rose & Parfitt, 2007). However, the promotion of autonomy is a central pillar of self-determination theory and its benefit is considered universal. It was, therefore, one of the main goals of the intervention. This underscores the perils of developing interventions and applying them on a large scale without pilot testing.

Furthermore, the intervention was delivered by non-specialists (mainly graduate students in exercise science, who lacked background in psychology, and students in psychology who lacked background in exercise science), following minimal training (two-day seminar), and through relatively minimal contact (only seven times in 8 months, on average, with only one of these lasting for an hour). As many experts on depression have noted, encouraging depressed adults to exercise is extremely challenging (Salmon, 2001; Seime & Vickers, 2006). Therefore, the decision to use a previously untested intervention and deliver it through non-specialists with minimal qualifications and experience shows that the magnitude of the challenge at hand was clearly

underestimated. No one would seriously consider delivering a psychological intervention or prescribing medications to a depressed population through anyone other than a highly trained professional with established credentials. So the assumption that a minimally trained non-specialist, guided only by a 25-page training manual, could conceivably make depressed patients exercise seems to indicate a lack of appreciation for the complexity of exercise motivation in a clinical context. Remarkably, not a single reference in the manual was made to the dysfunctional cognitions that are the hallmarks of depression and nothing in the manual gives any indication that this intervention was specifically developed for a population with this particular psychopathology.

As Salmon (2001) aptly described it, convincing a depressed patient to exercise is essentially tantamount to delivering effective psychotherapy, both in terms of difficulty and, to a large extent, in terms of technique: "Exercise in such patients is likely to depend on persuasive or therapeutic maneuvers of the kind that are integral to conventional psychological treatment. That is, the institution of exercise habits could be the *evidence* rather than the basis of successful treatment" (p. 38, emphasis in the original). To understand why anyone would think otherwise, one must understand that exercise is not taught in medical or clinical psychological curricula, a situation that often results in the erroneous assumption that exercise is "simple" (Connaughton, Weiler, & Connaughton, 2001; Faulkner & Biddle, 2001; Weiler, Chew, Coombs, Hamer, & Stamatakis, 2012).

The researchers emphasized that the "intervention increased physical activity" (Chalder et al., 2012, p. 5), with the implication being that depression scores were not reduced despite the fact that physical activity was increased. However, from a design standpoint, the intervention was deeply problematic. Because the patients in the "usual case" condition also received encouragement to join NHS exercise prescription schemes that are available throughout the country, the percentage of those in the usual care group who did at least 1,000 MET-minutes per week increased from 26.8% at baseline to 43.4% (by 16.6%) at 4 months. By comparison, the percentage for the intervention group increased from 24.7% to 51.5% (by 26.8%, only about 10% more than usual care). This difference was not statistically significant ($p = 0.08$) and, arguably, clinically negligible. This is crucially important because the "four month follow-up was chosen as the primary outcome endpoint" (p. 2), namely the point at which the outcomes of self-reported depression symptoms and antidepressant use were assessed. It should be clear, however, that at the 4-month time point (i.e., the "primary outcome endpoint"), the intervention had failed to produce a statistically significant (let alone a clinically meaningful) difference in physical activity levels. From an experimental standpoint, since the aim of the study, as stated, was to "evaluate . . . whether physical activity can be an effective treatment for depression" (Baxter et al., 2010, p. 6), once the "manipulation check" indicated that the intervention had failed, carrying on with the analyses makes no sense.

The 4-month time point was chosen because, according to the researchers, it "represented the stage in the intervention period at which [they] expected to observe the largest effect" (p. 2). However, the opposite was the case; not only was physical activity not significantly different between the usual care and the intervention groups at the 4-month time point, but their difference was, in fact, the smallest that was recorded and less than half of the largest difference (at 4 months, 51.5% versus 43.4%, or 8.1%; at 8 months, 63.2% versus 49.4%, or 13.8%; at 12 months, 57.7 versus 40.4, or 17.3%). By opting to compare the groups on the primary outcomes at the 4-month time point, the researchers chose to minimize the strength of the treatment and, by doing so, the statistical power of the analysis.

Furthermore, "there was no evidence that the difference between the groups changed over the duration of the study (time by treatment interaction, $p = 0.71$)" (Chalder et al., 2012, p. 5).

Even at the point of the largest effect of the intervention, which was actually 4 months after the intervention had ended (i.e., at 12 months post-randomization), there was only a 17% difference between the usual care and the intervention group in the percentage of those reaching the criterion of 1000 MET-minutes per week (at 12 months post-randomization). In design terms, there was severe cross-contamination across the experimental groups, since throughout the study 40%–50% of the participants in the usual care group were physically active at levels believed to be sufficient to effect change. Thus, by simultaneously providing encouragement for physical activity to the usual care group, the researchers, once again, chose to minimize the strength of the treatment and, by doing so, reduce the statistical power of the analysis.

Finally, both the measurement of physical activity and the measurement of depression (i.e., the two main variables of interest in the study) were problematic. Physical activity was measured through a 7-day recall diary. This method is known to be prone to a variety of biases and is among the least dependable ways to assess physical activity that are available today. The Beck Depression Inventory II, which was the sole measure of depression, was previously criticized by one of the coauthors of the Chalder et al. (2012) study as a measure that "is difficult to interpret clinically" (Lawlor & Hopker, 2001, p. 6). In spite of its popularity, the Beck Depression Inventory suffers from controversial factorial validity, easy susceptibility of scores to momentary changes in environmental conditions, and relatively poor discriminant validity against anxiety (Richter, Werner, Heerline, Kraus, & Sauer, 1998). For these reasons, sole reliance on the Beck Depression Inventory to operationalize the central outcome variable of the study was a questionable decision. It has become common practice in RCTs to corroborate the data based on this instrument with a different assessment method, namely a standardized clinician interview and rating scale (usually the Hamilton Rating Scale for Depression).

As the previous paragraphs demonstrate, the study by Chalder et al. (2012) had severe methodological flaws that make its findings irrelevant to the originally stated purpose of the study (i.e., to "evaluate . . . whether physical activity can be an effective treatment for depression"). In that sense, the statements in the press release were entirely misleading. However, it is crucial to point out that, by using the criteria for evaluating study quality that Lawlor and Hopker (2001) applied in their meta-analysis, the study by Chalder et al. (2012) would have been judged as being of high quality. Specifically, the authors noted that "treatment allocation, concealed from the study researchers using an automated telephone system, was administered remotely and employed a computer generated code" (p. 2). Furthermore, "the primary comparative analyses were conducted using an intention to treat approach" (p. 3). Finally, even though "owing to the nature of the intervention, none of the participants, general practices, clinicians, or researchers performing the outcome assessments could be blinded to treatment allocation" (p. 2), the use of "a self completion questionnaire" to assess depression was believed to "eliminate any observer bias" (p. 2). There is, therefore, little doubt that this study will appear in future systematic reviews and meta-analyses as being not only relevant to the question of the effect of physical activity on depression but also of high methodological quality. Thus, its null effect will be considered (perhaps with a high weight due to the large sample size) in the calculation of pooled standardized mean differences. This underscores the need for informed, balanced, and critical, as opposed to mechanical and superficial, evaluations of study quality. More broadly, the media coverage of the Chalder et al. study underscores the new challenges that patients and clinicians must face in their quest to make sense of the evidence on physical activity and mental health.

## Recapitulation

Mental health problems exact a significant toll on society, both in terms of disability and reduced quality of life and in terms of economic cost. Standard forms of treatment, namely psychotherapy and pharmacotherapy, help a lot of people suffering with mental health challenges. However, they show limited effectiveness for many individuals, they come with a high price for individuals and health care systems, and they do have a range of undesirable side-effects. In this context, physical activity offers a promising alternative form of prevention and treatment. It is safe, inexpensive, and demonstrably effective for a wide range of mental health problems. For these reasons, researchers and clinicians in psychology, psychiatry, primary care, and public health should be informed about the latest research discoveries on the relationship between physical activity and mental health from credible sources. Unfortunately, achieving this goal has been difficult and remains extremely challenging. The literature is plagued by contradictory statements and, due to the technical complexities involved, it is becoming increasingly difficult, especially for non-specialists, to parse the crucial details and make sense of the evidence. As more national clinical practice guidelines recommend the use of physical activity as a treatment option, the potential for bias in the published literature is bound to increase, further obfuscating the evidence. The ambitious goal of this handbook, authored by an international group of experts, is to summarize the current evidence in a comprehensive, balanced, and accessible manner, shedding light on controversial issues, highlighting the progress that has been made, and outlining an agenda for future research.

## References


Ahlskog, J.E., Geda, Y.E., Graff-Radford, N.R., & Petersen, R.C. (2011). Physical exercise as a preventive or disease-modifying treatment of dementia and brain aging. *Mayo Clinic Proceedings*, *86*, 876–884.

Alzheimer's Association (2010). 2010 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *6*, 158–194.

Andlin-Sobocki, P., Jonsson, B., Wittchen, H.U., & Olesen, J. (2005). Cost of disorders of the brain in Europe. *European Journal of Neurology*, *12* (Suppl 1), 1–27.

Arroll, B., Elley, C.R., Fishman, T., Goodyear-Smith, F.A., Kenealy, T., Blashki, G., Macgillivray, S. (2009). Antidepressants versus placebo for depression in primary care. *Cochrane Database of Systematic Reviews*, *2009* (3), CD007954.

Arroll, B., Macgillivray, S., Ogston, S., Reid, I., Sullivan, F., Williams, B., & Crombie, I. (2005). Efficacy and tolerability of tricyclic antidepressants and SSRIs compared with placebo for treatment of depression in primary care: A meta-analysis. *Annals of Family Medicine*, *3*, 449–456.

Babyak, M., Blumenthal, J.A., Herman, S., Khatri, P., Doraiswamy, M., Moore, K., . . . Krishnan, K.R. (2000). Exercise treatment for major depression: maintenance of therapeutic benefit at 10 months. *Psychosomatic Medicine*, *62*, 633–638.

Balon, R. (2006). SSRI-associated sexual dysfunction. *American Journal of Psychiatry*, *163*, 1504–1509.

Balshem, H., Helfand, M., Schünemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., . . . Guyatt, G.H. (2011). GRADE guidelines: Rating the quality of evidence. *Journal of Clinical Epidemiology*, *64*, 401–406.

Barlow, D.H. (2010). Negative effects from psychological treatments: A perspective. *American Psychologist*, *65*, 13–20.

Baskin, T.W., Callen Tierney, S., Minami, T., & Wampold, B.E. (2003). Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology*, *71*, 973–979.

Baxter, H., Winder, R., Chalder, M., Wright, C., Sherlock, S., Haase, A., . . . Lewis, G. (2010). Physical activity as a treatment for depression: The TREAD randomised trial protocol. *Trials*, *11*, 105.

Beck, C.A., Patten, S.B., Williams, J.V.A., Wang, J.L., Currie, S.R., Maxwell, C.J., & El-Guebaly, N. (2005). Antidepressant utilization in Canada. *Social Psychiatry and Psychiatric Epidemiology*, *40*, 799–807.

Beck, C.A., Williams, J.V.A., Wang, J.L., Kassam, A., El-Guebaly, N., Currie, S.R., Maxwell, C.J., & Patten, S.B. (2005). Psychotropic medication use in Canada. *Canadian Journal of Psychiatry*, *50*, 605–613.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Stroup, D.F. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*, *276*, 637–639.

Berk, M. (2007). Should we be targeting exercise as a routine mental health intervention? *Acta Neuropsychiatrica*, *19*, 217–218.

Biddle, S.J.H., Fox, K.R., Boutcher, S.H., & Faulkner, G.E. (2000). The way forward for physical activity and the promotion of psychological well-being. In S.J.H. Biddle, K.R. Fox, & S.H. Boutcher (Eds.), *Physical activity and psychological well-being* (pp. 154–168). London: Routledge.

Blanchflower, B.G., & Oswald, A.J. (2011). *Antidepressants and age*. Bonn, Germany: Institute for the Study of Labor.

Blumenthal, J.A., Babyak, M.A., Doraiswamy, P.M., Watkins, L., Hoffman, B.M., Barbour, K.A., . . . Sherwood, A. (2007). Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosomatic Medicine*, *69*, 587–596.

Blumenthal, J.A., & Ong, L. (2009). Commentary on "Exercise and depression" (Mead et al., 2009): And the verdict is . . . *Mental Health and Physical Activity*, *2*, 97–99.

Brosse, A.L., Sheets, E.S., Lett, H.S., & Blumenthal, J.A. (2002). Exercise and the treatment of clinical depression in adults: Recent findings and future directions. *Sports Medicine*, *32*, 741-760.

Bull, S.A., Hu, X.H., Hunkeler, E.M., Lee, J.Y., Ming, E.E., Markson, L.E., & Fireman, B. (2002). Discontinuation of use and switching of antidepressants: Influence of patient-physician communication. *Journal of the American Medical Association*, *288*, 1403–1409.

Callaghan, P. (2004). Exercise: A neglected intervention in mental health care? *Journal of Psychiatric and Mental Health Nursing*, *11*, 476–483.

Canadian Mental Health Association (2004). *Meeting the mental health needs of the people of Canada: A submission to the House of Commons Standing Committee on Finance*. Ottawa, ON: Author.

Chalder, M., Wiles, N.J., Campbell, J., Hollinghurst, S.P., Haase, A.M., Taylor, A.H., . . . Lewis, G. (2012). Facilitated physical activity as a treatment for depressed adults: Randomised controlled trial. *British Medical Journal*, *344*, e2758.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Colton, C.W., & Manderscheid, R.W. (2006). Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. *Preventing Chronic Disease*, *3* (2), 1–14.

Connaughton, A.V., Weiler, R.M., & Connaughton, D.P. (2001). Graduating medical students' exercise prescription competence as perceived by deans and directors of medical education in the United States: Implications for healthy people 2010. *Public Health Reports*, *116*, 226–234.

Cooper, C., Balamurali, T.B., & Livingston, G. (2007). A systematic review of the prevalence and covariates of anxiety in caregivers of people with dementia. *International Psychogeriatrics*, *19*, 175–195.

Cuijpers, P. (2005). Depressive disorders in caregivers of dementia patients: A systematic review. *Aging & Mental Health*, *9*, 325–330.

Daley, A.J. (2002). Exercise therapy and mental health in clinical populations: Is exercise therapy a worthwhile intervention? *Advances in Psychiatric Treatment*, *8*, 262–270.

Daviglus, M.L., Bell, C.C., Berrettini, W., Bowen, P.E., Connolly, E.S. Jr., Cox, N.J. . . . Maurizio Trevisan (2010). National Institutes of Health state-of-the-science conference statement: Preventing Alzheimer disease and cognitive decline. *Annals of Internal Medicine*, *153*, 176–181.

de Abajo, F.J., & García-Rodríguez, L.A. (2008). Risk of upper gastrointestinal tract bleeding associated with selective serotonin reuptake inhibitors and venlafaxine therapy: Interaction with nonsteroidal anti-inflammatory drugs and effect of acid-suppressing agents. *Archives of General Psychiatry*, *65*, 795–803.

Deshmukh, R., & Franco, K. (2003). Managing weight gain as a side effect of antidepressant therapy. *Cleveland Clinic Journal of Medicine*, *70*, 614–623.

Donaghy, M.E. (2007). Exercise can seriously improve your mental health: Fact or fiction? *Advances in Physiotherapy*, *9*, 76–88.

Dunn, A.L., Trivedi, M.H., Kampert, J.B., Clark, G.G., & Chambliss, H.O. (2005). Exercise treatment for depression efficacy and dose response. *American Journal of Preventive Medicine*, *28*, 1–8.

Ekkekakis, P. (2008). The genetic tidal wave finally reached our shores: Will it be the catalyst for a critical overhaul of the way we think and do science? *Mental Health and Physical Activity*, *1*, 47–52.

Ekkekakis, P., & Backhouse, S.H. (2009). Exercise and psychological well-being. In R. Maughan (Ed.), *Olympic textbook of science in sport* (pp. 251–271). Hoboken, NJ: Wiley-Blackwell.

Ensrud, K.E., Blackwell, T.L., Ancoli-Israel, S., Redline, S., Yaffe, K., Diem, S., . . . Stone, K.L. (2006). Use of selective serotonin reuptake inhibitors and sleep disturbances in community-dwelling older women. *Journal of the American Geriatrics Society*, *54*, 1508–1515.

Etters, L., Goodall, D., & Harrison, B.E. (2008). Caregiver burden among dementia patient caregivers: A review of the literature. *Journal of the American Academy of Nurse Practitioners*, *20*, 423–428.

Eysenck, H.J. (1952). The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, *16*, 319–324.

Faulkner, G., & Biddle, S. (2001). Exercise and mental health: It's just not psychology! *Journal of Sports Sciences*, *19*, 433–444.

Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., . . . Alzheimer's Disease International (2005). Global prevalence of dementia: A Delphi consensus study. *Lancet*, *366*, 2112–2117.

Flicker, L., Liu-Ambrose, T., & Kramer, A.F. (2011). Why so negative about preventing cognitive decline and dementia? The jury has already come to the verdict for physical activity and smoking cessation. *British Journal of Sports Medicine*, *45*, 465–467.

Folkins, C.H., & Sime, W.E. (1981). Physical fitness training and mental health. *American Psychologist*, *36*, 373–389.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R.C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *Journal of the American Medical Association*, *303*, 47–53.

Gartlehner, G., Hansen, R.A., Carey, T.S., Lohr, K.N., Gaynes, B.N., & Randolph, L.C. (2005). Discontinuation rates for selective serotonin reuptake inhibitors and other second-generation antidepressants in outpatients with major depressive disorder: A systematic review and meta-analysis. *International Clinical Psychopharmacology*, *20*, 59–69.

Gelenberg, A.J., Freeman, M.P., Markowitz, J.C., Rosenbaum, J.F., Thase, M.E., Trivedi, M.H., & Van Rhoads, R.S. (2010). *Practice guideline for the treatment of patients with major depressive disorder* (3rd ed.). Arlington, VA: American Psychiatric Association.

Goethe, J.W., Woolley, S.B., Cardoni, A.A., Woznicki, B.A., & Piez, D.A. (2007). Selective serotonin reuptake inhibitor discontinuation: Side effects and other factors that influence medication adherence. *Journal of Clinical Psychopharmacology*, *27*, 451–458.

González, H.M., Vega, W.A., Williams, D.R., Tarraf, W., West, B.T., Neighbors, H.W. (2010). Depression care in the United States: Too little for too few. *Archives of General Psychiatry*, *67*, 37–46.

Greenberg, P.E., Kessler, R.C., Birnbaum, H.G., Leong, S.A., Lowe, S.W., Berglund, P.A., & Corey-Lisle, P.K. (2003). The economic burden of depression in the United States: How did it change between 1990 and 2000? *Journal of Clinical Psychiatry*, *64*, 1465–1475.

Grissom, R.J. (1996). The magical number .7 ± .2: Meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology*, *64*, 973–982.

Guyatt, G.H., Cook, D.J., Sackett, D.L., Eckman, M., & Pauker, S. (1998). Grades of recommendation for antithrombotic agents. *Chest*, *114*, 441S–444S.

Haase, A.M., Taylor, A.H., Fox, K.R., Thorp, H., & Lewis, G. (2010). Rationale and development of the physical activity counselling intervention for a pragmatic TRial of Exercise and Depression in the UK (TREAD-UK). *Mental Health and Physical Activity*, *3*, 85–91.

Haskell, W.L., Lee, I.M., Pate, R.R., Powell, K.E., Blair, S.N., Franklin, B.A., . . . Bauman, A. (2007). Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine and Science in Sports and Exercise*, *39*, 1423–1434.

Healy, D. (2003). Lines of evidence on the risks of suicide with selective serotonin reuptake inhibitors. *Psychotherapy and Psychosomatics*, *72*, 71–79.

Healy, D., & Cattell, D. (2003). Interface between authorship, industry and science in the domain of therapeutics. *British Journal of Psychiatry*, *183*, 22–27.

Health Canada (2002). *A report on mental illnesses in Canada*. Ottawa, ON: Author.

Herbert, J.D., & Gaudiano, B.A. (2005). Moving from empirically supported treatment lists to practice guidelines in psychotherapy: The role of the placebo concept. *Journal of Clinical Psychology*, *61*, 893–908.

Hughes, J.R. (1984). Psychological effects of habitual aerobic exercise: A critical review. *Preventive Medicine*, *13*, 66–78.

Hunot, V.M., Horne, R., Leese, M.N., & Churchill, R.C. (2007). A cohort study of adherence to antidepressants in primary care: The influence of antidepressant concerns and treatment preferences. *Primary Care Companion to the Journal of Clinical Psychiatry*, *9*, 91–99.

Ilyas, S., & Moncrieff, J. (2012). Trends in prescriptions and costs of drugs for mental disorders in England, 1998–2010. *British Journal of Psychiatry*, *200*, 393–398.

Insel, K.R. (2008). Assessing the economic costs of serious mental illness. *American Journal of Psychiatry*, *165*, 663–665.

Ioannidis, J.P.A. (2008). Effectiveness of antidepressants: An evidence myth constructed from a thousand randomized trials? *Philosophy, Ethics, and Humanities in Medicine*, *3*, 14.

Jick, H., Kaye, J.A., & Jick, S.S. (2004). Antidepressants and the risk of suicidal behaviors. *Journal of the American Medical Association*, *292*, 338–343.

Juurlink, D.N., Mamdani, M.M., Kopp, A., & Redelmeier, D.A. (2006). The risk of suicide with selective serotonin reuptake inhibitors in the elderly. *American Journal of Psychiatry*, *163*, 813–821.

Kannel, W.B., & Sorlie, P. (1979). Some health benefits of physical activity: The Framingham study. *Archives of Internal Medicine*, *139*, 857–861.

Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27.

Kennedy, S.H., & Rizvi, S. (2009). Sexual dysfunction, depression, and the impact of antidepressants. *Journal of Clinical Psychopharmacology*, *29*, 157–164.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., & Walters, E.E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602.

Kessler, R.C., Heeringa, S., Lakoma, M.D., Petukhova, M., Rupp, A.E., Schoenbaum, M., . . . Zaslavsky, A.M. (2008). Individual and societal effects of mental disorders on earnings in the United States: Results from the National Comorbidity Survey Replication. *American Journal of Psychiatry*, *165*, 703–711.

Kirsch, I. (2005). Placebo psychotherapy: Synonym or oxymoron? *Journal of Clinical Psychology*, *61*, 791–803.

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., & Johnson, B.T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine*, *5*, e45.

Kisely, S., Cox, M., Campbell, L.A., Cooke, C., & Gardner, D. (2009). An epidemiologic study of psychotropic medication and obesity-related chronic illnesses in older psychiatric patients. *Canadian Journal of Psychiatry*, *54*, 269–274.

Krogh, J., Nordentoft, M., Sterne, J.A.C., & Lawlor, D.A. (2011). The effect of exercise in clinically depressed adults: Systematic review and meta-analysis of randomized controlled trials. *Journal of Clinical Psychiatry*, *72*, 529–538.

Kuhn, T.S. (1962/1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.

Lambert, M.J. (2005). Early response in psychotherapy: Further evidence for the importance of common factors rather than "placebo effects." *Journal of Clinical Psychology*, *61*, 855–869.

Landman, A., Cortese, D.K., & Glantz, S. (2008). Tobacco industry sociological programs to influence public beliefs about smoking. *Social Science & Medicine*, *66*, 970–981.

Lawlor, D.A., & Hopker, S.W. (2001). The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised control trials. *British Medical Journal*, *322*, 1–8.

Levit, K.R., Kassed, C.A., Coffey, R.M., Mark, T.L., McKusick, D.R., King, E., . . . Stranges, E. (2008). *Projections of national expenditures for mental health services and substance abuse treatment, 2004–2014* (SAMHSA Publication No. SMA 08-4326). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Lim, K.-L., Jacobs, P., Ohinmaa, A., Schopflocher, D., & Dewa, C.S. (2008). A new population-based measure of the economic burden of mental illness in Canada. *Chronic Diseases in Canada*, *28*, 92–98.

Lipsey, M.W., & Wilson, D.B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209.

Lockhart, P., & Guthrie, B. (2011). Trends in primary care antidepressant prescribing 1995–2007: A longitudinal population database analysis. *British Journal of General Practice*, *61*, e565–e572.

MacCoun, R.J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, *49*, 259–287.

Mark, T.L., Levit, K.R., Coffey, R.M., McKusick, D.R., Harwood, H.J., King, E.C., . . . Ryan, K. (2007). *National expenditures for mental health services and substance abuse treatment, 1993–2003* (SAMHSA Publication No. SMA 07–4227). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Mathers, C.D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Medicine*, *3* (11), e442.

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). *Paying the price: The cost of mental health care in England to 2026*. London: King's Fund.

McGoey, L. (2010). Profitable failure: Antidepressant drugs and the triumph of flawed experiments. *History of the Human Sciences*, *23*, 58–78.

Mead, G.E., Morley, W., Campbell, P., Greig, C.A., McMurdo, M., & Lawlor, D.A. (2009). Exercise for depression. *Cochrane Database of Systematic Reviews* (3), CD004366.

Moher, D., Jones, A., & Lepage, L. (2001). Use of the CONSORT statement and quality of reports of randomized trials: A comparative before-and-after evaluation. *Journal of the American Medical Association*, *285*, 1992–1995.

Mojtabai, R., & Olfson, M. (2008). National trends in psychotherapy by office-based psychiatrists. *Archives of General Psychiatry*, *65*, 962–970.

Moore, M., Yuen, H.M., Dunn, N., Mullee, M.A., Maskell, J., & Kendrick, T. (2009). Explaining the rise in antidepressant prescribing: A descriptive study using the general practice research database. *British Medical Journal*, *339*, b3999.

Morgan, W.P. (1981). Psychological benefits of physical activity. In F.J. Nagle & H.J. Montoye (Eds.), *Exercise in health and disease* (pp. 299–314). Springfield, IL: Charles C. Thomas.

Murray, C.J., & Lopez, A.D. (1997). Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet*, *349*, 1498–1504.

National Collaborating Centre for Mental Health and National Institute for Health and Clinical Excellence (2010). *Depression: The treatment and management of depression in adults* (updated edition). Leicester and London: British Psychological Society and Royal College of Psychiatrists.

Nutt, D.J., & Sharpe, M. (2008). Uncritical positive regard? Issues in the efficacy and safety of psychotherapy. *Journal of Psychopharmacology*, *22*, 3–6.

Oksbjerg Dalton, S., Johansen, C., Mellemkjær, L., Nørgård, B., Toft Sørensen, H., & Olsen, J.H. (2003). Use of selective serotonin reuptake inhibitors and risk of upper gastrointestinal tract bleeding: A population-based cohort study. *Archives of Internal Medicine*, *163*, 59–64.

Olfson, M., & Marcus, S.C. (2009). National patterns in antidepressant medication treatment. *Archives of General Psychiatry*, *66*, 848–856.

Olfson, M., & Marcus, S.C. (2010). National trends in outpatient psychotherapy. *American Journal of Psychiatry*, *167*, 1456–1463.

Olfson, M., Marcus, S.C., Druss, B., & Pincus, H.A. (2002). National trends in the use of outpatient psychotherapy. *American Journal of Psychiatry*, *159*, 1914–1920.

Olfson, M., Marcus, S.C., Tedeschi, M., & Wan, G.J. (2006). Continuity of antidepressant treatment for adults with depression in the United States. *American Journal of Psychiatry*, *163*, 101–108.

Otto, M.W., Church, T.S., Craft, L.L., Greer, T.L., Smits, J.A.J., & Trivedi, M.H. (2007). Exercise for mood and anxiety disorders. *Journal of Clinical Psychiatry*, *68*, 669–676.

Parloff, M.B. (1986). Placebo controls in psychotherapy research: A sine qua non or a placebo for research problems? *Journal of Consulting and Clinical Psychology*, *54*, 79–87.

Patten, S.B., Esposito, E., & Carter, B. (2007). Reasons for antidepressant prescriptions in Canada. *Pharmacoepidemiology and Drug Safety*, *16*, 746–752.

Phillips, W.T., Kiernan, M., & King, A.C. (2003). Physical activity as a nonpharmacological treatment for depression: A review. *Complementary Health Practice Review*, *8*, 139–152.

Plassman, B.L., Williams, J.W. Jr., Burke, J.R., Holsinger, T., & Benjamin, S. (2010). Systematic review: Factors associated with risk for and possible prevention of cognitive decline in later life. *Annals of Internal Medicine*, *153*, 182–193.

Posternak, M.A., & Zimmerman, M. (2007). Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: Meta-analysis. *British Journal of Psychiatry*, *190*, 287–292.

Prioleau, L., Murdock, M., & Brody, N. (1983). An analysis of psychotherapy versus placebo studies. *Behavioral and Brain Sciences*, *6*, 275–310.

Rethorst, C.D., Wipfli B.M., & Landers, D.M. (2009). The antidepressive effects of exercise: A meta-analysis of randomized trials. *Sports Medicine*, *39*, 491–511.

Richter, P., Werner, J., Heerlein, A., Kraus, A., & Sauer, H. (1998). On the validity of the Beck Depression Inventory: A review. *Psychopathology*, *31*, 160–168.

Rimer, J., Dwan, K., Lawlor, D.A., Greig, C.A., McMurdo, M., Morley, W., & Mead, G.E. (2012). Exercise for depression. *Cochrane Database of Systematic Reviews* (7), CD004366.

Rose, E.A., & Parfitt, G. (2007). A quantitative analysis and qualitative explanation of the individual differences in affective responses to prescribed and self-selected exercise intensities. *Journal of Sport & Exercise Psychology*, *29*, 281–309.

Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: A unifying theory. *Clinical Psychology Review*, *21*, 33–61.

Schulz, K.F., Chalmers, I., Hayes, R.J., & Altman, D.G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association*, *273*, 408–412.

Schulz, R., Boerner, K., Shear, K., Zhang, S., & Gitlin, L.N. (2006). Predictors of complicated grief among dementia caregivers: A prospective study of bereavement. *American Journal of Geriatric Psychiatry*, *14*, 650–658.

Schwartz, T.L., Nihalani, N., Jindal, S., Virk, S., & Jones, N. (2004). Psychiatric medication-induced obesity: A review. *Obesity Reviews*, *5*, 115–121.

Scottish Intercollegiate Guidelines Network (2010). *Non-pharmaceutical management of depression in adults: A national clinical guideline*. Edinburgh: Author.

Seime, R.J., & Vickers, K.S. (2006). The challenges of treating depression with exercise: From evidence to practice. *Clinical Psychology: Science and Practice*, *13*, 194–197.

Senger, H.L. (1987). The "placebo" effect of psychotherapy: A moose in the rabbit stew. *American Journal of Psychotherapy*, *41*, 68–81.

Serretti, A., & Chiesa, A. (2009). Treatment-emergent sexual dysfunction related to antidepressants: A meta-analysis. *Journal of Clinical Psychopharmacology*, *29*, 259–266.

Smith, E.A. (2006). "It's interesting how few people die from smoking": Tobacco industry efforts to minimize risk and discredit health promotion. *European Journal of Public Health*, *17*, 162–170.

Smith, M.L., & Glass, G.V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752–760.

Snowden, M., Steinman, L., Mochan, K., Grodstein, F., Prohaska, T.R., Thurman, D.J., . . . Anderson, L.A. (2011). Effect of exercise on cognitive performance in community-dwelling older adults: review of intervention trials and recommendations for public health practice and research. *Journal of the American Geriatrics Society*, *59*, 704–716.

Soni, A. (2009). *The five most costly conditions, 1996 and 2006: Estimates for the U.S. civilian noninstitutionalized population* (Statistical Brief 248). Rockville, MD: Agency for Healthcare Research and Quality.

Stephens, T., & Joubert, N. (2001). The economic burden of mental health problems in Canada. *Chronic Diseases in Canada*, *22*, 18–23.

Taylor, A.H., & Faulkner, G. (2008). Inaugural editorial. *Mental Health and Physical Activity*, *1*, 1–8.

Tkachuk, G.A., & Martin, G.L. (1999). Exercise therapy for patients with psychiatric disorders: Research and clinical implications. *Professional Psychology: Research and Practice*, *30*, 275–282.

Turner, E.H., Matthews, A.M., Linardatos, E., Tell, R.A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, *358*, 252–260.

Turner, M.S., May, D.B., Arthur, R.R., & Xiong, G.L. (2007). Clinical impact of selective serotonin reuptake inhibitors therapy with bleeding risks. *Journal of Internal Medicine*, *261*, 205–213.

United States Department of Health and Human Services (1996). *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: Author, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

United States Department of Health and Human Services (1999). *Mental health: A report of the Surgeon General*. Rockville, MD: Author, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.

United States National Center for Health Statistics (2010). *Health, United States, 2009: With special feature on medical technology*. Hyattsville, MD: United States Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Üstün, T.B., Ayuso-Mateos, J.L., Chatterji, S., Mathers, C., & Murray, C.J. (2004). Global burden of depressive disorders in the year 2000. *British Journal of Psychiatry*, *184*, 386–392.

Wampold, B.E., Mondin, G.W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin*, *122*, 203–215.

Wang, P.S., Berglund, P., Olfson, M., Pincus, H.A., Wells, K.B., & Kessler, R.C. (2005). Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 603–613.

Weiler, R., Chew, S., Coombs, N., Hamer, M., & Stamatakis, E. (2012). Physical activity education in the undergraduate curricula of all UK medical schools: Are tomorrow's doctors equipped to follow clinical guidelines? *British Journal of Sports Medicine, 46,* 1024–1026.

WHO World Mental Health Survey Consortium (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Journal of the American Medical Association*, *291*, 2581–2590.

Wimo, A., Jönsson, L., & Winblad, B. (2006). An estimate of the worldwide prevalence and direct costs of dementia in 2003. *Dementia and Geriatric Cognitive Disorders*, *21*, 175–181.

Wimo, A., Winblad, B., & Jönsson, L. (2010). The worldwide societal costs of dementia: Estimates for 2009. *Alzheimer's & Dementia*, *6*, 98–103.

Wipfli, B.M., Rethorst, C.D., & Landers, D.M. (2008). The anxiolytic effects of exercise: A meta-analysis of randomized trials and dose-response analysis. *Journal of Sport & Exercise Psychology*, *30*, 392–410.

Wittchen, H.U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., . . . Steinhausen, H.C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*, 655–679.

World Health Organization (2003). *Investing in mental health*. Geneva, Switzerland: Author.