Chapter 6

Physical Activity Status of People with Mental Disorders compared to the General Population: a Systematic Review of Longitudinal Cohort Studies

This following article in review is based on the this chapter;

Suetani S, Stubbs B, McGrath J, and Scott J. Physical activity status of people with mental disorders compared to the general population: a systematic review of longitudinal cohort studies. *In Submission.*

Overview

In this Chapter, we investigated the Thesis Question Three; *what is the longitudinal association between mental disorders and PA status?* There has been limited examination of the bidirectional association between PA and mental disorders using longitudinal studies. We examined if (a) people with lower PA have an increased risk of subsequent mental disorders (compared to those with higher PA); and, more relevant to the thesis, (b) people with mental disorders have reduced subsequent PA (compared to those without mental disorders).

We conducted a systematic review of longitudinal studies examining mental disorders and PA. Longitudinal studies were population-based. Mental disorders were defined by ICD or DSM. The findings were described in a narrative summary.

We identified eighteen studies. The majority (sixteen) examined affective disorders and PA. Three studies found a reduced risk of subsequent affective disorders in those with increased PA, five reported mixed results varying by sex and amount of PA, and five found no association. One study reported an increased risk of reduced PA in those with prior affective disorders but three other studies examining the relationship found mixed results. Of four studies examining anxiety disorders and PA, one study found that those with prior anxiety disorders had reduced subsequent PA. Among three studies examining the reciprocal relationship, no association was identified between PA and subsequent anxiety disorders. None of the included studies found association between PA and subsequent substance use or psychotic disorders.

There is a lack of consistent evidence linking PA to be either a risk factor or consequence of mental disorders.

Introduction

PA is generally defined as “any bodily movement produced by skeletal muscles that results in energy expenditure” (Caspersen et al., 1985). It has been well-established as an important modifiable risk factor for the cardio-metabolic related mortality in the general population (Lee et al., 2012). Compared to the general population, people with mental disorders engage in significantly less PA (Stubbs et al., 2016a, Schuch et al., 2017), and PA interventions have demonstrated benefits in both psychological and physical well-beings of people with mental disorders (Rosenbaum et al., 2016a, Rosenbaum et al., 2014). Therefore, PA is increasingly being recognized as a potential treatment target to reduce cardio-metabolic burden and associated differential mortality in those with mental disorders (Lederman et al., 2017).

While much of the existing evidence for the efficacy of PA among people with mental disorders consists of findings from randomized controlled trials (Firth et al., 2015), they are often limited by factors such as selection bias (e.g. participants who complete the study are more likely to be motivated to be physically active than non-participants) and short study duration. Cohort studies can mitigate some of these limitations and may facilitate examination of “real world effectiveness” of PA. While a few cross-sectional studies (Suetani et al., 2016b, Suetani et al., 2017b) have examined the relationship between PA and mental disorders, they are less able to explore the temporal relationship between the variables of interest. In terms of the longitudinal association, Mammen and Faulkner (2013) found that 25 out of 30 longitudinal studies identified demonstrated that increased PA was associated with a reduced risk of subsequent depression. Roshanaei-Moghaddam and colleagues (2009), on the other hand, examined the effect of baseline depression on subsequent PA. Eight out of eleven studies included found that depression at baseline was associated with subsequent reduction in PA. More recently, a comprehensive systematic review by Schuch *et al* (2018) examined 49 longitudinal studies consisting of 266,935 participants to find that those with high levels of PA had lower odds of developing depression (OR 0.83; 95% CI 0.79 - 0.88), and these effects were seen among youth, adults, and elderly participants. However, previous reviews were based a broad range of measures of depression – some were

symptom scales, and others diagnostic instruments. This has implications for clinical interpretation of the findings and casts doubt on the validity of pooling such widely disparate measures of depression (Mammen and Faulkner, 2013).

To the best of our knowledge, there have been no reviews that systematically examined; (i) the bidirectional relationship between mental disorders and PA, and (ii) the association between mental disorders and PA in people with mental disorders other than depressive disorder (e.g. anxiety disorders, substance use disorders, or psychotic disorders). Thus, this systematic review examines the following research questions: (i) Do people with lower PA have an increased risk of subsequent mental disorders – including affective disorders, anxiety disorders, substance use disorders, and psychotic disorders - compared to those with higher PA?, and, to answer one of the Thesis Question: (ii) Do people with mental disorders have reduced subsequent PA (compared to those without mental disorders)? In order to optimize the clinical utility of the analyses, this review focused on traditional diagnostic criteria for mental disorders.

Methods

A systematic review was conducted according to Preferred reporting Items for Systematic Reviews and Meta-Analyses (PRSMA) guideline (Moher et al., 2009) and Meta-Analysis of Observational Studies in Epidemiology (MOOSE) guideline (Stroup et al., 2000). This study was prospectively registered in advance with PROSPERO (registration ID CRD42017071737). The following databases were searched for English-language, original research articles published in peer-review journals from inception to September 2017: Medline, PubMed, EMBASE, PsychInfo, CINAHL, and Web of Science. In addition, references from articles identified as well as several systematic reviews in the field were examined to identify any other article eligible for the current review. The following search algorithm was used: (“exercise” OR “physical activity”) AND (“schizophrenia” OR “psychosis” OR “depression” OR “bipolar” OR "serious mental illness" OR "severe mental illness" OR "anxiety" OR "substance use disorder" OR "substance dependence" OR "alcohol use disorder" OR "alcohol dependence" OR "stimulant use disorder" OR "stimulant dependence"

OR "common mental illness" OR "common mental disorder") AND (“longitudinal” OR “observational”). The inclusion criteria of the current review were: (a) longitudinal observational study with either prospective or retrospective design of at least twelve months (or one year) of follow up; (b) study populations include people with mental disorders (defined as those who meet the diagnostic criteria, either ICD (any version) or DSM (any version), for CMD (defined as affective disorders, anxiety disorders or substance use disorders) and SMD (defined as psychotic disorders including schizophrenia spectrum disorders and bipolar disorder) as well as people without mental disorders of interest (“the general population”); (c) presence of mental disorder diagnoses or self-reported or objectively measured PA at the study initial follow-up point; and (d) outcome measures of either change in PA over time in those with mental disorders v general population or incidence of mental disorders. The exclusion criteria were; (a) experimental studies, intervention studies, reviews, meta-analyses, cross-sectional studies, or case-control studies; and (b) studies without general population (e.g. psychiatric inpatient cohort).

Data Extraction

Titles and abstracts of the articles were reviewed to identify studies that met the eligibility criteria. Eligibility assessment was then performed by two authors (SS and BS) independently. All articles that met eligibility criteria were coded. In keeping with previous studies (Stubbs et al., 2016c, Stubbs et al., 2016a), the following characteristics were extracted from each study when available; (a) study description (including author, publication year, location, study design, follow-up period, sample numbers, loss to follow-up, age, gender, PA measures, and mental disorder measures), and (b) study findings (effect size metrics, 95% CI, and confounders adjusted for). The study findings were examined and summarized according to consistency in direction and significance of the results. Two independent researchers extracted the data (SS and BS) and disagreements were resolved by consensus.

Methodological Quality Appraisal

The assessment of risk of bias in each study was evaluated using Newcastle–Ottawa Scale (NOS) (Wells) (Table 6.1). The NOS is a method recommended by the Cochrane Non-Randomized Studies Methods Working Group to evaluate the quality of the study. Points are assigned based on the selection process of cohorts (0–4 points), the comparability of the cohorts (0–2 points) and the identification of the exposures and the outcomes of research participants (0–3 points). A score of at least 7 out of 9 was defined as high quality. Two reviewers (SS and BS) independently assessed the methodological quality of each study with disagreements resolved by discussion.

Additional Analyses

The following sensitivity analyses were pre-specified in the review protocol; sub-analyses by (i) sex, (ii) age group, (iii) self-reported v objective PA measurement, (iv) country, and (v) persistence or recovery of mental health symptoms.

Statistical Analysis

Descriptive statistics were used to present a summary of the findings. As studies included were not sufficiently homogeneous in terms of exposure, comparator and outcomes, meta-analysis of studies was not conducted in the current review.

Results

The search of Medline, PubMed, EMBASE, PsychINFO, CINAHL, and Web of Science databases identified a total of 8,407 potential papers (Figure 6.1). After removing duplicates, 4,665 papers remained. An additional five papers were located through reference lists from other systematic reviews on the topic. Of these 4,670 papers, 4,647 papers were discarded after reviewing the titles and abstracts. The full texts of remaining 23 papers were examined. Six papers required further discussion before consensus was reached. Three corresponding authors were contacted via email for further data and/or clarification. All three authors responded to our correspondence and two provided us with further data. Five of these papers were

excluded due to; (a) PA not being reported as an independent variable (n = 1), (b) studies being case control in design (n =2), and (c) neither ICD nor DSM criteria were used for diagnoses (n = 2). A total of 18 studies were identified for inclusion in the review.

A summary of included studies is presented in Table 6.2. Of 18 studies included, three studies examined the association between baseline mental disorders and subsequent PA (Patten et al., 2009, Naicker et al., 2013, Hiles et al., 2017), and only one study (Jerstad et al., 2010) reported on the bi-directional relationship between the variables of interest. The other 14 studies examined the association between baseline PA and subsequent mental disorders. The number of participants ranged from 496 (Jerstad et al., 2010) to 10,625 (Mikkelsen et al., 2010) and follow up periods ranged from 15 months (Henchoz et al., 2014) to 26 years (Mikkelsen et al., 2010).

With respect to type of mental disorder, the majority of the included studies (11 out of 18) examined depressive disorders (Wang et al., 2011, Tanaka et al., 2011, Strawbridge et al., 2002, Mikkelsen et al., 2010, McKercher et al., 2014, Jerstad et al., 2010, Jacka et al., 2011, Colman et al., 2014, Cabello et al., 2017, Patten et al., 2009, Naicker et al., 2013), while one examined both depressive disorders and anxiety disorders (Hiles et al., 2017) and another (Weyerer, 1992) examined depressive and bipolar disorders**.** Two studies examined the relationship between PA and alcohol use disorder (Henchoz et al., 2014, Ejsing et al., 2015). Two studies assessed affective disorders, anxiety disorders, and substance use disorders within the same samples (Ten Have et al., 2011, Strohle et al., 2007). One study (Suetani et al., 2017a) examined affective disorders, anxiety disorders, substance use disorders as well as psychotic disorders within the same cohort group. The majority of the studies (13 out of 18) utilized DSM criteria (Ten Have et al., 2011, Tanaka et al., 2011, Suetani et al., 2017a, Strohle et al., 2007, Strawbridge et al., 2002, Patten et al., 2009, Naicker et al., 2013, McKercher et al., 2014, Jerstad et al., 2010, Jacka et al., 2011, Hiles et al., 2017, Henchoz et al., 2014, Colman et al., 2014), while four used ICD criteria (Weyerer, 1992, Mikkelsen et al., 2010, Ejsing et al., 2015, Cabello

et al., 2017). One study did not specify the diagnostic criteria used but was included because the study utilized CIDI questionnaire to derive its diagnoses (Wang et al., 2011).

In contrast to mental disorder diagnoses, there was little consistency in how PA was measured and categorized. None of the included study used objective PA measurement tools. Only six studies used validated questionnaires like IPAQ (The IPAQ Group, 2005), with remaining twelve studies relying on isolated self-reported PA items, often from just one question (e.g. “Over the past 12 months, how often did you participate in sports or exercising?” (Henchoz et al., 2014)). These questionnaires often used exercise (defined as a subtype of PA that is repetitive and structured, and has a specific intention of improving or maintaining fitness (Caspersen et al., 1985)) as a proxy for the overall PA. The wide range of measurement tools utilized to estimate the PA led to inconsistent PA categorizations (i.e. different studies used the same terminologies (e.g. “physically active”) to describe a wide range of different intensities/types/frequencies of PA), thus compromising our ability to compare and synthesize included studies in a systematic manner.

Quality assessment scores for included studies are summarized in Table 6.3. Utilizing the NOS modified specifically for the current study, 15 out of 18 included studies were assessed as having high quality (i.e. score of at least 7 out of 9). The mean score was 7.1 with the lowest score of 5 and the highest score of 8.

Primary Analyses

The association between PA and mental disorders were examined for affective disorders, anxiety disorders, substance use disorders, and psychotic disorders.

Affective Disorders

Sixteen out of 18 studies examined the association between affective disorders and PA. Most (13/16) focused on depressive disorder, but three examined wider diagnostic entities to include dysthymia (Suetani et al., 2017a) and bipolar disorders (Ten Have et al., 2011, Strohle et al., 2007). Of these 16 studies, 13 (Weyerer, 1992, Wang et al., 2011, Ten Have et al., 2011, Tanaka et al., 2011, Suetani et al., 2017a, Strohle et al., 2007, Strawbridge et al., 2002, Mikkelsen et al., 2010, McKercher et al., 2014, Jerstad et al., 2010, Jacka et al., 2011, Colman et al., 2014, Cabello et al., 2017) examined the relationship between PA at baseline and subsequent affective disorders; nine (Jerstad et al., 2010, Mikkelsen et al., 2010, Cabello et al., 2017, Strawbridge et al., 2002, Tanaka et al., 2011, Wang et al., 2011, Weyerer, 1992, Strohle et al., 2007, Ten Have et al., 2011) of which examined for incident (new onset) affective disorders, whilst four studies (Colman et al., 2014, Jacka et al., 2011, McKercher et al., 2014, Suetani et al., 2017a) did not measure the baseline affective disorder status. One (Cabello et al., 2017) also examined the effect of persistent depression on PA. Four studies (Hiles et al., 2017, Naicker et al., 2013, Patten et al., 2009, Jerstad et al., 2010) examined the impact of affective disorder at baseline on subsequent PA.

In terms of incident affective disorders (i.e. new cases of affective disorders at the follow up), four out of nine studies (Cabello et al., 2017, Wang et al., 2011, Weyerer, 1992, Strohle et al., 2007) found no association between PA at baseline and subsequent onset of affective disorders. Two studies (Jerstad et al., 2010, Strawbridge et al., 2002) found significant association between higher PA and a reduced likelihood of incident affective disorders. The remaining three studies had mixed results; Mikkelsen *et al* (2010) found that women with low PA had an increased likelihood of incident depression (compared to those with high PA). There were no significant associations between women with moderate PA compared to high PA, or in men. PA in this study was defined based on the response to the question: “If you should state your LEISURE TIME PHYSICAL ACTIVITY including transport to and from work during the last year, in which group do you think you should be placed (one answer only)?” Participants were given the following choices; (a) almost entirely inactive or light PA less than two hours per week (defined in this study as low PA), (b) light PA for two to four hours per week (defined in this study as

moderate PA), (c) light PA for more than four hours per week or more vigorous activity for two to four hours per week, or, (d) more vigorous PA for more than four hours per week or regular heavy exercise, or competitive sports several times per week (the last two categories combined together to form the high PA group). Tanaka *et al* (2011), on the other hand, found that men who reported to "never" engage in PA had an increased risk of incident depression compared to those who reported to "often, sometimes" engage in PA. There was no significant association between the two variables found in women in this study. Ten Have *et al* (2011) found that those who engaged in between one to three hours per week exercising, but not in those who exercised for more than four hours per week, had a reduced risk of subsequent mood disorders (compared to those who engaged in no exercise). PA in this study was estimated based on the response to the question: "How many hours per week have you engaged in physical exercise/sport lately?"

Four studies examined the association of baseline PA and subsequent affective disorders without assessing for baseline affective disorder diagnoses, making it impossible to determine if these were incident cases or not. One study (Colman et al., 2014) found no association between PA and affective disorders in adolescents, while another study (Jacka et al., 2011) found that lower PA was associated with an increased risk of subsequent depression. Using the response to the question: "How often did you exercise or play sports in the last week?", Suetani *et al* (2017a) found that compared to adolescents who engaged in frequent PA (more than four days), those with no PA engagement ("Not at all") had an increased risk of subsequent affective disorders. However, there was no association between the two variables when the infrequent PA engagement (one to three days) was compared to the frequent PA engagement group. McKerracher *et al* (2014) used a self-reported past-week frequency and duration of school and extracurricular sport and exercise at baseline (aged 9 to 15), and the long form IPAQ (Craig et al., 2003) at follow-up (approximately 20 years later) to categorize participants into four PA groups; (i) decreasing, (ii) increasing, (iii) persistently active, and (iv) persistently inactive. The study found that men in the increasing or the persistently active PA groups, but not those in the decreasing group, had a reduced risk of subsequent depression

(compared to men in the persistently inactive group). No difference among the four groups were seen in women.

Of specific interest to this thesis, four studies examined the effect of affective disorders on subsequent PA. Jerstad *et al* (2010) found that adolescent girls with depression were less likely to engage in PA (compared to the non-depressed cohort) six years later. Two studies (Hiles et al., 2017, Naicker et al., 2013) found no association between depression at baseline and subsequent PA. One study (Patten et al., 2009) found that participants with baseline depression who were active (defined in this study as the total estimated energy expenditure greater than 1.5 kcal/kg per day) had a greater likelihood of transitioning into being physically inactive (compared to those without depression). However, among those who were physically inactive at baseline, depression was not associated with the transition into being physically active at follow up.

Overall, these studies reveal an inconsistent pattern of associations between PA and affective disorders.

Anxiety Disorders

Four studies (Hiles et al., 2017, Strohle et al., 2007, Suetani et al., 2017a, Ten Have et al., 2011) examined the association between anxiety disorders and PA. Of these, two studies (Strohle et al., 2007, Ten Have et al., 2011) examined incident anxiety disorders and one study (Suetani et al., 2017a) did not examine baseline anxiety disorders. Only one study explored the impact of anxiety disorders on subsequent PA (Hiles et al., 2017).

Of two studies examining incident anxiety disorders, Strohle *et al* (2007) found that compared to those engaged in no PA (defined in this study as less than once a month or no exercise at all), participants who were engaged in regular PA (defined in this study as daily and several times a week), but not in non-regular PA (defined in

this study as one to four times a month), showed a reduced risk of subsequent anxiety disorders. On the other hand, Ten Have *et al* (2011) found that participants who engaged in between one to three hours per week of exercise, but not those who engaged in more than four hours per week of exercise, showed a reduced risk of subsequent anxiety disorders (compared to those engaged in no exercise). The study by Suetani *et al* (2017a) found no association between PA and anxiety disorders (though this study did not control for the baseline diagnosis). Overall, these studies suggest no consistent association between PA and subsequent anxiety disorders.

The only study that examined the association between the baseline anxiety disorders and subsequent PA found that anxiety disorders diagnosis was associated with reduced subsequent PA (Hiles et al., 2017).

Substance Use Disorders

Five studies examined the association between substance use disorders and PA (Henchoz et al., 2014, Ejsing et al., 2015, Strohle et al., 2007, Suetani et al., 2017a, Ten Have et al., 2011). Two studies explored alcohol use disorder, and the remaining three explored substance use disorders in general. One study (Ten Have et al., 2011) found no association between PA and subsequent substance use disorders. Three studies found mixed results. Ejsing *et al* (2015) estimated the PA levels using the response to the question; “If you should state your LEISURE TIME PHYSICAL ACTIVITY including transport to and from work during the last year, in which group do you think you should be placed (one answer only)?” Although this is the same question as the one used in the study by Mikkelsen *et al* (2010), Ejsing *et al* used different terminologies to define each category as follows; (a) almost entirely inactive or light PA less than two hours per week (defined in this study as a sedentary level), (b) light PA for two to four hours per week (defined in this study as a low level), (c) light PA for more than four hours per week or more vigorous PA for two to four hours per week, and (d) more vigorous PA for more than four hours per week or regular heavy exercise or competitive sports several times per week (the last two categories combined together to form a moderate/high level in this study).

They found that participants with a sedentary level, but not those with a low PA level, had an increased risk of subsequent alcohol use disorder (compared with those with a moderate/high PA level). This finding was found in both men and women. Likewise, Henchoz *et al* (2014) utilised a single question to estimate PA: “Over the past 12 months, how often did you participate in sports or exercising?” with response choices; “never”, “a few times a year”, “1–3 times per month”, “at least once per week”, or “almost every day”. They found that participants engaging in no sports or exercise (“never”) , but not those with lower engagement (“a few times a year”, “1–3 times per month”, and “at least once per week”), had an increased risk of subsequent alcohol use disorders (compared to those who engaged in sports or exercise “almost every day”). Strohle *et al* (Strohle et al., 2007) found that non-regular PA, but not regular PA, was associated with a reduced risk of subsequent substance use disorders (compared to those with no PA). Finally, Suetani *et al* (2017a) found that adolescents who engaged in infrequent PA, but not those who engaged in no PA, had a reduced risk of substance use disorders (compared to adolescents who engaged in frequent PA). Overall, there was no consistent association across all studies of PA and subsequent substance use disorders.

None of the included studies examined the association between the baseline substance use disorders and subsequent PA.

Psychotic Disorders

Only one study examined the association between PA and psychotic disorders. Suetani *et al* (2017a) did not find any association between the baseline PA and subsequent psychotic disorders.

Additional Analyses

The following sub-analyses were planned in the prospective protocol; (i) sex, (ii) age group, (iii) self-reported v objective PA measurements, (iv) geographical location, and (v) persistence or recovery of mental health symptoms. As no included study

utilised objective PA measurement, the sub-analysis (iii) was not done. Likewise, all but one study (combined data from Ghana, India, Mexico and Russia) was conducted in developed countries (Cabello et al., 2017), with the majority of studies coming from either North America (Naicker et al., 2013, Patten et al., 2009, Jerstad et al., 2010, Colman et al., 2014, Strawbridge et al., 2002, Wang et al., 2011) or Europe (Hiles et al., 2017, Mikkelsen et al., 2010, Henchoz et al., 2014, Weyerer, 1992, Ejsing et al., 2015, Strohle et al., 2007, Ten Have et al., 2011), with three coming from Australia (Jacka et al., 2011, McKercher et al., 2014, Suetani et al., 2017a), and one from Japan (Tanaka et al., 2011). Thus, the sub-analysis (iv) was not conducted.

Sub-analysis by Sex

For affective disorders, seven studies reported results separately for each sex (Jerstad et al., 2010, Mikkelsen et al., 2010, McKercher et al., 2014, Tanaka et al., 2011, Wang et al., 2011, Weyerer, 1992, Suetani et al., 2017a). All reported the association between PA and subsequent affective disorders while one also explored the reciprocal relationship. Of seven that examined the association in women, one study showed association between higher PA and a reduced risk of subsequent affective disorders (Jerstad et al., 2010), one showed mixed results (Mikkelsen et al., 2010), and five remaining studies showed no associations (McKercher et al., 2014, Tanaka et al., 2011, Wang et al., 2011, Weyerer, 1992, Suetani et al., 2017a). Similarly, in the six studies that examined at the association in men, one showed that reduced PA was associated with an increased risk of subsequent mood disorders (Tanaka et al., 2011), one showed mixed results (McKercher et al., 2014), and four remaining studies showed no association (Mikkelsen et al., 2010, Wang et al., 2011, Weyerer, 1992, Suetani et al., 2017a). Only one study explored the impact of mental disorders and subsequent PA. In this study of adolescent girls, those with depression at baseline showed lower subsequent PA (Jerstad et al., 2010).

Only one study reported results by each sex for anxiety disorders. In an Australian study of young people, Suetani *et al* found that boys who had no PA, but not those with infrequent PA, had an increased risk of subsequent anxiety disorders

(compared to those who engaged in frequent PA). The same study found no association between PA and subsequent anxiety disorders in girls (Suetani et al., 2017a).

For substance use disorders, three studies reported results by each sex. One (Suetani et al., 2017a) found infrequent PA, but not no PA, was associated with a reduced risk of subsequent substance use disorders (compared to those with frequent PA) in girls. The same study found no association between PA and substance use disorder in boys. In contrast, the study by Ejsing *et al* (Ejsing et al., 2015) found that compared to those with a moderate/high PA level, both women and men with a sedentary level, but not those with a low PA level, had an increased risk of subsequent substance use disorders. Further, a study conducted in Switzerland of young male military recruits (Henchoz et al., 2014) found that reporting to engage in no sports or exercise, but not lower engagement (i.e. “a few times a year”, “1–3 times per month”, and “at least once per week”), was associated with an increased risk of subsequent alcohol use disorders in men (compared to those who engaged in sports or exercise “almost every day”).

Finally, Suetani *et al* (2017a) examined the relationship between PA and the subsequent risk of psychotic disorders in boys and girls separately. There was no association between these variables in either sex.

Sub-analysis by Age Group

Nine studies consisted of the general adult cohorts (Hiles et al., 2017, Patten et al., 2009, Mikkelsen et al., 2010, Cabello et al., 2017, Jacka et al., 2011, Wang et al., 2011, Weyerer, 1992, Ejsing et al., 2015, Ten Have et al., 2011). Six of these explored PA at baseline and subsequent affective disorders. One study found that lower PA was associated with an increased risk of subsequent depression (Jacka et al., 2011), three found no association (Cabello et al., 2017, Wang et al., 2011, Weyerer, 1992), and two found mixed results (Mikkelsen et al., 2010, Ten Have et al., 2011). Two studies examined affective disorders at baseline and subsequent PA.

One study found no association (Hiles et al., 2017) while the other found mixed results (Patten et al., 2009). In terms of anxiety disorders, one study found mixed results when examining the relationship between PA and subsequent diagnoses (Ten Have et al., 2011) while another study found that those with baseline anxiety disorders were more likely to have reduced subsequent PA (Hiles et al., 2017). Finally, when examining the association between baseline PA and subsequent substance use disorder in this age group, one study found mixed results (Ejsing et al., 2015) and another no association (Ten Have et al., 2011).

Seven studies (Naicker et al., 2013, Jerstad et al., 2010, Henchoz et al., 2014, Colman et al., 2014, McKercher et al., 2014, Strohle et al., 2007, Suetani et al., 2017a) focused on young people under the age of 20 at the onset of the study. Of these, five studies examined baseline PA and subsequent affective disorders. One found that adolescent girls with reduced PA were more likely to develop incident depressive disorder subsequently (Jerstad et al., 2010), and two studies each showed either no association (Colman et al., 2014, Strohle et al., 2007) or mixed results (McKercher et al., 2014, Suetani et al., 2017a). Two studies explored the reciprocal relationship between the baseline affective disorders and subsequent PA. While Jerstad *et al* (2010) found that adolescent girls with depression were less likely to engage in subsequent PA, Naicker *et al* (2013) found no association between the two variables. In addition, two studies found mixed results for baseline PA and subsequent anxiety disorders (Strohle et al., 2007, Suetani et al., 2017a), three studies found mixed results for baseline PA and subsequent substance use disorders (Strohle et al., 2007, Suetani et al., 2017a, Henchoz et al., 2014), and one found no association between baseline PA and subsequent psychotic disorders (Suetani et al., 2017a) in this age group.

Two studies examined older adults with one consisting of those over the age 40 (Tanaka et al., 2011), and another with those over the age 50 (Strawbridge et al., 2002). Tanaka *et al* (2011) found an increased risk of subsequent affective disorders in men with lower PA, but this relationship was not identified in women. Strawbridge *et al* (2002) reported that the higher PA at baseline was associated with a reduced

risk of subsequent affective disorders. Neither of the two studies examined any other diagnostic group.

Sub-analysis by Persistence or Recovery of Mental Health Symptoms

Cabello *et al* (2017) examined the sub-group who had persistent depression (defined as those who met the criteria for depression at both baseline and follow up). Using the IPAQ PA categories (The IPAQ Group, 2005), they found that participants with low PA, but not those with moderate PA, were more likely to have persistent depression (compared to those with high PA).

Ten Have *et al* (2011) examined if PA (defined in this study as hours per week engaged in physical exercise/sport recently) was associated with the likelihood of recovery from various mental disorders (affective disorders, anxiety disorders, and substance use disorder). PA was not associated with recovery from affective disorders or substance use disorders, but there was evidence to suggest that those who had engaged in exercise/sports were more likely to recover from anxiety disorders (compared to those who had not engaged in exercise/sports recently).

Discussion

Summary of Evidence

We would like to draw attention to three key findings. First, apart from depression, there has been relatively little research examining the association between PA and mental disorders using clinical diagnostic criteria. Second, there is a wide range of different methods used to measure and classify PA, thus making it difficult to compare one study from another. Third, the findings from this Chapter - to our knowledge, the first systematic review to examine the association between PA and mental disorder diagnoses using longitudinal observation studies - suggest there is no convincing data to support the hypothesis that PA influences the risk of subsequent mental disorders, nor that mental disorders impact on subsequent PA.

The 18 studies included 15,338 participants in total. Within these studies, only two found consistent association between lower PA and a reduced risk of subsequent mental disorders. One study found the bidirectional association (i.e. greater PA was associated with a reduced likelihood of subsequent major depression and major depression predicted lower subsequent PA engagement). Eleven studies found mixed results (i.e. no consistency in direction and significance of the findings). Four studies found no association between the variables of interest.

Previously, Mammen and Faulkner (2013) reported that 25 out of 30 longitudinal studies identified in their review showed the association between greater PA and a reduced subsequent risk of developing depression. However, only five (Jerstad et al., 2010, Mikkelsen et al., 2010, Strawbridge et al., 2002, Wang et al., 2011, Weyerer, 1992) of the 30 studies included in their study met the inclusion criteria for the current review due to this review having a narrower definition of mental disorders. Similarly, Hoare *et al* (2014) have identified six longitudinal studies that examined the association between PA and subsequent depressive symptoms among adolescents. Even though all six studies found the association between greater PA and a reduced subsequent risk of depressive symptoms with one study also demonstrating the opposite direction of association, only one out of the six studies included (Jerstad et al., 2010) met the inclusion criteria for the current review.

There has been little attention given to the association between the PA and subsequent anxiety, substance use, and psychotic disorders. In the current review, four studies (Hiles et al., 2017, Strohle et al., 2007, Suetani et al., 2017a, Ten Have et al., 2011) examined the association between PA and anxiety disorders. Five studies (Strohle et al., 2007, Suetani et al., 2017a, Ten Have et al., 2011, Ejsing et al., 2015, Henchoz et al., 2014) examined the association between PA and subsequent substance use disorders and only one study examined the relationship between PA and psychotic disorders (Suetani et al., 2017a). Overall, there was no consistent association among these studies.

Strengths and Limitations

Strengths of the current analysis included the particular focus on participants with more robust definitions of mental disorders, thus allowing the findings to have more clinical utility compared to previous reviews. It also focused on data from longitudinal studies, facilitating a better exploration of the temporal association between PA and mental disorders.

The main limitation of the current analysis is the inconsistency in measurement and definition of PA in the included studies and absence of objective measurement of PA. As PA is often unstructured and varies in both intensity and duration as well as type, measuring and assessing PA is inherently complex (Knowles, 2017). More accurate and reliable PA measurements would allow us to explore more fine-grained aspects of the relationship such as the dose-response relationship, as well as the impact of duration, intensity and pattern of PA (Dowd et al., 2018). Further, the risk of under-reporting with self-report measurements may be particularly high in people with mental disorders. For example, using the data from the UK Biobank, Firth *et al* (2017) recently demonstrated that self-reported PA questionnaire failed to capture the significant reduction in PA engagement demonstrated with accelerometers. Given that all the studies included in this review relied on self-reports rather than objective measurements, and the proportion of participants focused were those with mental disorders, the risk of underreporting due to both recall/reporting biases and methodological effectiveness of the tools used is likely to be substantial.

The majority of included studies measured mental disorder diagnoses at both time points to allow for the estimate of incident, or change in mental disorder diagnosis status. This allowed the current review to investigate if PA was associated with the change in the outcome (mental disorder diagnosis status), rather than just examining the association between the two variables at different time points. However, few included studies accounted for the change in PA status using a consistent measurement tool over the same period of time, which limited our ability to examine the change in the PA status in the same manner. Further, the current review only examined one mental disorder diagnosis at a time. Given comorbidity of CMD is very

common (e.g. one large Australian study found that nearly half (48.5%) of people with CMD had more than one CMD diagnosis (Suetani et al., 2017b)), it would be useful for future studies in this field to consider the impact of CMD comorbidity on PA status.

Another important limitation is the fact that the current review only included observational studies. While most of the included studies controlled for potential confounders such as age, sex and BMI, only few of them considered other factors like and smoking status, other medical conditions and socioeconomic status, and none of them included factors like CRF (the ability of the circulatory, respiratory, and muscular systems to supply oxygen during sustained PA). There may be other unmeasured confounding factors that may influence the association between the two variables in any given observational study. Given these limitations, even though the current review restricted its search to longitudinal studies only, it is unable to comment on causative factors between mental disorders and PA. This may, at least in part, explain why the results of the current review are not in line with the accumulating base of research evidence from interventional studies that indicate that PA is beneficial to people with mental disorders (Rosenbaum et al., 2014, Rosenbaum et al., 2016a).

In addition, consistent with previous reviews, the current analysis found that the majority of samples identified came from developed countries. There were also only two studies that explored older participants while seven studies consisted of young people. While this may reflect the fact that most mental disorders have its onset in youth (Kessler et al., 2007a), it also limits the generalisability of the current findings in older people.

Future Direction

The current Chapter highlights the need for researchers in this field to explore an appropriate and consistent tool to measure PA in large cohort studies with more robust designs. While objective measurements such as accelerometers may have

more favourable features in terms of validity, reliability, and sensitivity when measuring PA, this needs to be weighed against practical issues such as cost, feasibility, acceptability, and tolerance, especially if the study requires the participants to wear these devices for a considerable amount of time. Furthermore, given how rapidly the technology is advancing, most measurement devices that are used at the baseline data collection may be out of date by the follow up point in longitudinal studies with long durations. Thus, subjective validated measurement tools such as self-report questionnaire still have an important role (Bauman et al., 2009). It may also be useful to have a tool that has specifically considered needs of people with mental disorders. Currently, there is an active exploration in this area as exemplified by the world wide evaluation of SIMPAQ (simple PA questionnaire) (Rosenbaum et al., 2016b). Alternatively, or in combination with measurements for PA status, another option is to use a measurement of CRF. In the general population, CRF predicts mortality better than PA, especially in the older cohort (Kujala, 2018). A proxy measure of CRF such as the six minute walk test may be relatively easy to perform (Laboratories, 2002), and the understanding of the importance of CRF in people with mental disorders is growing (Stubbs et al., 2016b, Vancampfort et al., 2015b, Vancampfort et al., 2015c). At the population level, measures of CRF may represent a more suitable modifiable risk factor and potential consequences of mental disorders.

PA is a complex behaviour with many different psychosocial factors such as social support influencing the opportunity to engage in PA across the life course (Bauman et al., 2012, Suetani et al., 2016a). Likewise, mental disorders have various known and unknown causative pathways affected by both biological and psychosocial factors (e.g. (Suetani et al., 2018b)). To account for such complexities, future studies should incorporate novel designs that may help strengthen the exploration of the causative natures of the relationship such as Mendelian comparison and sibling comparison on a large scale in a collaborative manner (Kujala, 2018, Wade et al., 2018).

Conclusion

At this stage, there is a lack of consistent evidence linking PA to be either a risk factor or consequence of mental disorders. The current review also found the need for a more consistent approach to measuring and defining PA, as well as incorporating novel approaches to account for inherent complexities of PA. Given the potential importance of PA at both the individual and population levels, this issue warrants ongoing attention.

Acknowledgements

We would like to thank the following researchers for providing additional data; Yves Henchoz from Lusanne University Hospital, Sarah Hiles from VU University Medical Center, and Arne Okkenhaug from Levanger Hospital.

Table 6.1: Newcastle-Ottawa Quality Assessment Scale for Cohort Studies criteria used for this review

Note: A study can be awarded a maximum of one star (*) for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection
1) Representativeness of the exposed cohort **a) truly representative of the average people with mental disorders in the community*** **b) somewhat representative of the average people with mental disorders in the community*** **c) selected group of users eg nurses, volunteers (particular age group or only one gender)** **d) no description of the derivation of the cohort**
2) Selection of the non-exposed cohort **a) drawn from the same community as the exposed cohort*** **b) drawn from a different source** **c) no description of the derivation of the non exposed cohort**
3) Ascertainment of exposure **a) objective measure (e.g. accelerometer)*** **b) structured interview or validated self-report* (e.g. IPAQ)** **c) written self report (non-validated questions)** **d) no description**
4) Demonstration that outcome of interest was not present at start of study **a) yes*** **b) no**
Comparability
1) Comparability of cohorts on the basis of the design or analysis **a) study controls for baseline mental disorder diagnosis or physical activity status*** **b) study controls for any additional factors (age, sex and BMI)***

Outcome
1) Assessment of outcome **a) independent blind assessment*** **b) record linkage*** **c) self report** **d) no description**
2) Was follow-up long enough for outcomes to occur **a) yes (5 years)*** **b) no**
3) Adequacy of follow up of cohorts **a) complete follow up - all subjects accounted for*** **b) subjects lost to follow up unlikely to introduce bias - small number lost - > 80% follow up, or description provided of those lost)*** **c) follow up rate < 79% and no description of those lost** **d) no statement**

Table 6.2: Summary of included studies evaluating longitudinal association of physical activity status and mental disorder diagnoses

Source	Age group at baseline and sex	Site	Number of participants at follow up	Follow up	PA measurement categories (Measurement used)	Mental disorder diagnoses (Measurement used)	Main findings	Association
Studies examining the association between baseline physical activity and later mental disorders								
Cabello et al 2017	18 and over Male and female	Ghana, India, Mexico, Russia	7,908	3 to 8 years	Low, moderate, high (IPAQ-SF in wave 0, GPAQ in wave 1)	ICD-10 depressive episode (CIDI)	Low, but not moderate, PA predicted higher risk of persistent depression No association between PA and incident depression	Mixed
Colman et al 2014	12 to 17 years Male and female	Canada	583	14 years	Physically active, moderately active, physically inactive (Energy index)	DSM-IV major depressive episode (CIDI)	No association between PA and depression	No
Ejsing et al 2015	20 and over Male and female (analysed separately)	Denmark	6,237	20.9 years	Sedentary level, low level, moderate/high level (One question self-report)	ICD-10 alcohol use disorder (National Patient Registry)	Sedentary, but not low, level was associated with subsequent alcohol use disorder in	Mixed

							both male and female	
Henchoz et al 2014	Mean age 19.95 Male only	Switzerland	4,846	15 months	Resisters, lapsers, adopters, maintainers (One question self-report)	DSM-IV alcohol dependence (self-report questionnaire)	Resisters and lapsers, but not adopters, showed increased risk of subsequent alcohol dependence	Mixed
Jacka et al 2011	20 to 97 years Male and female	Australia	2,152	Retrospective study	Low PA, high PA (Retrospective questionnaire)	DSM-IV depression (self-report questionnaire)	Low PA predicted higher risk of subsequent depression	Yes
McKercher et al 2014	9 to 15 years Male and female	Australia	1,630	20 years	Decreasing, increasing, persistently active, persistently inactive (Historical leisure activity questionnaire on baseline, IPAQ at follow up)	DSM-IV major depressive disorder or dysthymic disorder (CIDI)	No association between PA and subsequent depression in female Increasing, and persistently active PA associated with reduced risk of subsequent depression in male	Mixed

Mikkelsen et al 2010	20 to 93 years Male and female	Denmark	10,625	26 years	Low level, moderate level, high level (One question self-report)	ICD-10 depression (National Patient Registry)	In female, increased risk of future depression found in those in the low PA, but not moderate, group In male, no association found between PA and future depression	Mixed
Strawbridge et al 2002	50 and over Male and female	United States	1,947	5 years	PA scale ranging from 0 to 8 (Self-report questionnaire)	DSM-IV depression (CIDI)	Higher PA associated with a reduced risk of subsequent depression	Yes
Strohle et al 2007	14 to 24 years Male and female	Germany	2,548	4 years	Regular, non-regular, no (Self-report questionnaire)	DSM-IV anxiety disorders, mood disorders, substance use disorders (CIDI)	Reduced risk of subsequent anxiety disorders in those with regular PA Reduced risk of subsequent	Mixed

							substance use disorders in those with non-regular PA No association between mental disorders and PA status otherwise	
Suetani et al 2017	Mean age 13.9 Male and female	Australia	2,575	7 years	No, infrequent, frequent (Self-report questionnaire)	DSM-IV anxiety disorder, mood disorders, substance use disorders, psychotic disorders (CIDI)	Increased risk of subsequent mood disorders in those with no PA engagement Reduced risk of subsequent substance use disorders in those with infrequent PA engagement In male, increased risk of subsequent	Mixed

							anxiety disorders in no PA group In female, reduced risk of subsequent substance use disorders in those with infrequent PA engagement No other association between PA engagement and mental disorders	
Tanaka et al 2011	40 to 69 years Male and female	Japan	9,201	7 years	Often/sometimes, never (One question self-report)	DSM-IV major depressive episode (self-report questionnaire)	Increased risk of future depressive episode in those in never PA group in male No association between PA and future depressive episode in female	Mixed

ten Have et al 2011	18 to 64 years Male and female	Netherlands	4,796	3 years	0 hours per week spent on exercise, 1-3 hours per week, more than 4 hours per week (One question self-report)	DSM-III-R mood disorders, anxiety disorders, substance use disorders (CIDI)	Reduced risk of future mood disorders in those with 1-3h, but not 4h and more, spent on exercise Reduced risk of future anxiety disorders in those with 1-3h, but not 4h and more, spent on exercise No association between exercise and future substance use disorders	Mixed
Wang et al 2011	12 and over Male and female	Canada	Unclear	6 years	Active, inactive (Minnesota leisure-time PA questionnaire)	Depression (unclear which classification) (CIDI)	No association between PA and depression in either male or female	No

Weyerer 1992	15 and over Male and female	Germany	1,341	5 years	No, occasional, regular exercise (One question self-report)	ICD mood disorders (clinical interview scales)	No association between PA and future depression for total, male or female	No
Studies examining the association between baseline mental disorders and future physical activity								
Hiles et al 2017	18 to 65 years Male and female	Netherlands	1,994	6 years	MET minute per week (IPAQ)	DSM-IV anxiety disorder and DSM-IV depressive disorder (CIDI)	Current anxiety, current depression, current co-morbid, but not remitted disorder, were associated with reduced subsequent general PA	Mixed
Naicker et al 2013	16 to 17 years Male and female	Canada	681	10 years	Inactive, moderate/active (unclear)	DSM-IV major depressive episode (CIDI)	No association between adolescence depression and subsequent PA	No
Patten et al 2009	12 and over Male and female	Canada	8,422	10 years	Active, inactive (Self-report questionnaire)	DSM-IV major depressive episode (CIDI)	Major depressive episode predicted transition from active to	Mixed

							inactive PA but not from inactive to active PA	
Studies examining the bidirectional association between mental disorders and physical activity								
Jerstad et al 2010	11 to 15 years Female only	United States	496	6 years	Number of physical activities (Modified version of the Past Year Activity Scales)	DSM-IV depression (Schedule for affective disorder and schizophrenia for school aged children)	Greater PA associated with a reduced likelihood of subsequent major depression Major depression predicted lower subsequent PA engagement	Yes

PA, physical activity; IPAQ SF, international physical activity questionnaire short form; GPAQ, global physical activity questionnaire; ICD, international classification of diseases; CIDI, composite international diagnostic interview; DSM, diagnostic and statistical manual of mental disorders; MET, metabolic equivalent total

Table 6.3. Quality Assessment Scores

	Selection 1	Selection 2	Selection 3	Selection 4	Comparability 1	Outcome 1	Outcome 2	Outcome 3	Total
Cabello et al 2017	1	1	1	1	2	1	0	0	7
Coleman et al 2014	1	1	0	1	2	1	1	0	7
Ejsing et al 2015	1	1	0	1	2	1	1	1	8
Henchoz et al 2014	1	1	0	1	1	0	0	1	5
Hiles et al 2017	0	1	1	1	2	1	1	1	8
Jacka et al 2011	0	1	0	1	2	0	1	1	6
Jerstad et al 2010	0	1	1	1	2	1	1	1	8
McKercher et al 2014	1	1	1	0	2	0	1	1	7
Mikkelsen et al 2010	1	1	0	1	2	1	1	0	7
Naicker et al 2013	1	1	0	1	2	0	1	1	7
Patten et al 2009	1	1	1	0	2	1	1	0	7
Strawbridge et al 2002	1	1	0	1	2	1	1	1	8
Strohle et al 2007	1	1	0	1	2	1	0	1	7
Suetani et al 2017	1	1	0	0	2	1	1	1	7
Tanaka et al 2011	0	1	0	1	2	0	1	1	6

ten Have et al 2011	1	1	0	1	2	1	0	1	7
Wang et al 2011	1	1	1	1	2	1	1	0	8
Weyerer 1992	1	1	0	1	2	1	1	1	8

Figure 6.1: Search